

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
4 July 2002 (04.07.2002)

PCT

(10) International Publication Number  
WO 02/051992 A2

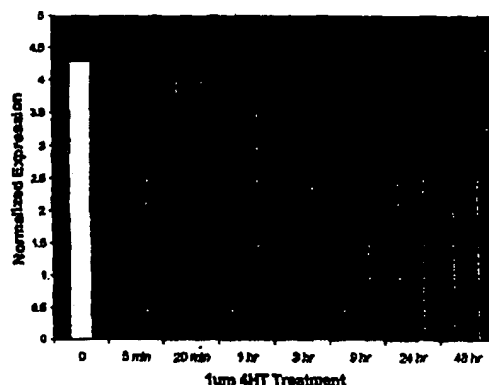
- (51) International Patent Classification<sup>7</sup>: C12N 9/00 (74) Agents: WILLIAMSON, Kelly, J. et al.; Alston & Bird LLP, Bank of America Plaza, Suite 4000, 101 South Tryon Street, Charlotte, NC 28280-4000 (US).
- (21) International Application Number: PCT/US01/50317
- (22) International Filing Date:  
20 December 2001 (20.12.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/257,736 22 December 2000 (22.12.2000) US  
10/025,950 19 December 2001 (19.12.2001) US
- (71) Applicant (for all designated States except US): MILLENIUM PHARMACEUTICALS, INC. [US/US]; 75 Sydney Street, Cambridge, MA 02139 (US).
- (72) Inventors; and  
(75) Inventors/Applicants (for US only): HUNTER, John, Joseph- [US/US]; 35 Harvard Street, Somerville, MA 02143 (US). KAPPELLER-LIBERMANN, Rosana [BR/US]; 86 Beacon Street, Chestnut Hill, MA 02467 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report

[Continued on next page]

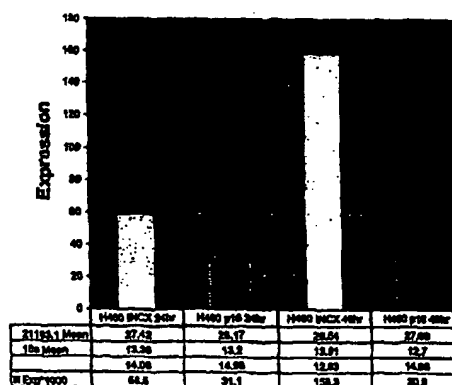
(54) Title: 21163; A NOVEL HUMAN PROLYL OLIGOPEPTIDASE AND USES THEREFOR

## Tumor Suppressor Induced Down-regulation of 21163 Expression

a) p53 induced down-regulation (TxP, MPG5)



b) p16 induced down-regulation (Taqman)



(57) Abstract: The invention provides isolated nucleic acid molecules, designated 21163 nucleic acid molecules, which encode novel prolyl oligopeptidase family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 21163 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 21163 gene has been introduced or disrupted. The invention still further provides isolated 21163 proteins, fusion proteins, antigenic peptides and anti-21163 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

Best Available Copy



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## 21163, A NOVEL HUMAN PROLYL OLIGOPEPTIDASE AND USES THEREFOR

### FIELD OF THE INVENTION

The present invention relates to a newly identified protein, 21163, a human prolyl oligopeptidase. In particular, the invention relates to 21163 prolyl oligopeptidase polypeptides and polynucleotides, methods of detecting the 21163 prolyl oligopeptidase polypeptides and polynucleotides, and methods of diagnosing and treating 21163 prolyl oligopeptidase-related disorders. Also provided are vectors, host cells, and recombinant methods for making and using the novel molecules.

### BACKGROUND OF THE INVENTION

10       The trypsin family of serine proteases is characterized by the presence of a conserved nucleophilic serine residue that is necessary for catalytic activity. A charge relay system provided by the catalytic triad consisting of a serine (nucleophile), histidine (base), and aspartate (electrophile) residue is the well-known catalytic mechanism of this family of proteases. The prolyl oligopeptidase family  
15       consists of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution.

      The prolyl oligopeptidase family includes prolyl endopeptidase (PE) which has been found in many human tissues including brain, skeletal muscle, and liver  
20       (Toide *et al.* (1998), *Reviews in the Neurosciences* 9:17-29), lymphocytes (Vanhoof *et al.* (1994) *Gene* 149:363-366), and T cells (Shirasawa *et al.* (1994) *J. Biochem.* 115:724-729). PE cleaves peptide bonds on the carboxyl-terminal side of prolyl residues. Dipeptidyl peptidase IV (DPP IV) has been shown to be expressed in a variety of human tissues including placenta, kidney, lung and liver (Abbott *et al.*  
25       (1994), *Immunogenetics* 40:331-338), and removes amino-terminal dipeptides sequentially from polypeptides having unsubstituted amino-termini provided that the penultimate residue is a proline. Acylamino-acid-releasing enzyme (acyl-peptide

hydrolase; ACPH) has been found in tissues such as porcine brain (Richards *et al.* (2000) *Mol. Pharmacol.* 58:557-583) and intestine (Giardina *et al.* (1999) *Biochimie* 81:1049-1055), bovine lens (Chongcharoen and Sharma (1998) *Biochem. Biophys. Res. Commun.* 247:136-144), and human erythrocytes, and catalyzes the hydrolysis of the amino-terminal peptide bond of an amino-acetylated protein to generate an amino-acetylated amino acid and a protein with a free amino-terminus.

DPP IV and PE are widely distributed in human tissues and body fluids and have been suggested to have a role in the regulation of intracellular protein turnover and in the degradation and processing of peptide hormones and neuropeptides (Welches *et al.* 1993) *Life Sciences* 52:1461-1480). For example, PE has been implicated in the pathogenesis of hypertension through renin-angiotensin regulation (Goossens *et al.* (1996) *Eur. J. Clin. Chem. Clin. Biochem.* 34:17-22). In addition, increased PE and in some cases DPP IV activity has been correlated with tumors of prostate, lung and sigmoid, whereas decreased levels of PE activity have been observed in sera of individuals suffering from HIV infection, malaria, prostate cancer, benign prostate hypertrophy, and erythroleukemia cells during DMSO-induced differentiation (Goossens *et al.* (1996) *Eur. J. Clin. Chem. Clin. Biochem.* 34:17-22; Sedo *et al.* (1991) *J. Cancer Res. Clin. Oncol.* 117:249-253); Tsukahara *et al.* (1991) *Int. J. Biochem.* 23:79-83). Extracellular DPP IV has been implicated in a variety of biological functions including T-cell activation, cell-to-cell adhesion, activation of signal transduction, cell matrix adherence, and in HIV infection (Abbott *et al.* (1994) *Immunogenetics* 40:331-338; David *et al.* (1993) *J. Biol. Chem.* 268:17247-17252).

PE plays an important role in the metabolism of proline-containing neuropeptides such as substance P (SP), arginine-vasopressin (AVP), and thyrotropin-releasing hormone (TRH) which are peptides capable of enhancing learning and memory (Toide *et al.* (1998) *Reviews in the Neurosciences* 9:17-29). There is also evidence that PE may participate in the symptomatology and/or etiology of Alzheimer's disease through proteolysis of beta-amyloid precursor protein leading to insoluble amyloid beta peptide (Barelli *et al.* (1999) *Biochem. Biophys. Res. Commun.* 57:657-661).

Furthermore, levels of PE are reduced in the neurodegenerative disorders, Lewy body dementia, Parkinson's disease, and Huntington's disease, indicating that

this enzyme may also play a role in these diseases (Mantle *et al.* (1996) *Clin. Chim. Acta.* 249:129-139. In addition, increased levels of serum PE activity have been correlated with stress-induced anxiety (Maes *et al.* (1998) *Psychoneuroendocrinology* 23:485-495) and psychotic conditions such as mania and schizophrenia (Maes *et al.* (1995) *Psychiatry Res.* 58:217-225). Decreased levels of serum PE activity could also play a role in the pathophysiology of major depression (Maes *et al.* (1995) *Psychiatry Res.* 58:217-225) and the aberrant pain perception and depressive symptoms in fibromyalgia (Maes *et al.* (1998) *Psychoneuroendocrinology* 23:485-495).

Similar to PE, ACPH has also been implicated in cognitive enhancement.

10 ACPH has been identified as a target of certain organophosphorus compounds with cognitive-enhancing effects, thus making it a potential target for cognitive-enhancing drugs (Richards *et al.* (2000) *Mol. Pharmacol.* 58:557-583). ACPH also appears to have a role in the regulation of cell growth, as the inhibition of ACPH leads to the inhibition of cell growth resulting in apoptosis (Yamaguchi *et al.* (1999) *Biochem.* 15 *Biophys. Res. Common.* 263:139-142).

It is well established that members of the prolyl oligopeptidase family of peptidases play critical roles in a variety of important cellular processes including the regulation of protein turnover; the processing and degradation of bioactive peptides including hormones and neuropeptides; cellular growth and proliferation,

20 differentiation, and apoptosis; T-cell activation; and cell-to-cell adhesion. As a result of these roles, the prolyl oligopeptidases are involved in such important diseases and disorders as anxiety, psychotic and depressive disorders, cognitive disorders, pain perception, neurodegenerative disorders, cancers, and infections such as HIV and malaria.

25 Accordingly, prolyl oligopeptidases are a major target for drug action and development. Therefore, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown prolyl oligopeptidases. The present invention advances the state of the art by providing a previously unidentified human prolyl oligopeptidase.

## SUMMARY OF THE INVENTION

The present invention is based, in part, on the discovery of a novel human prolyl oligopeptidase, referred to herein as "21163". The nucleotide sequence of a cDNA encoding 21163 is shown in SEQ ID NO:1, and the amino acid sequence of a 21163 polypeptide is shown in SEQ ID NO:2. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:3.

Accordingly, in one aspect the invention features a nucleic acid molecule which encodes a 21163 protein or polypeptide, e.g., a biologically active portion of the 21163 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides an isolated 21163 nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, wherein the nucleic acid encodes a full length 21163 protein or an active fragment thereof.

In a related aspect, the invention further provides nucleic acid constructs which include a 21163 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 21163 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing 21163 nucleic acid molecules and polypeptides.

In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 21163-encoding nucleic acids.

In still another related aspect, isolated nucleic acid molecules that are antisense to a 21163 encoding nucleic acid molecule are provided.

In another aspect, the invention features 21163 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 21163-mediated or 21163-related disorders. In another embodiment, the invention provides 21163 polypeptides having a 21163 activity. Preferred polypeptides are 21163 proteins including at least one prolyl oligopeptidase domain, and, preferably, having a 21163 activity, e.g., a 21163 activity as described herein.

In other embodiments, the invention provides 21163 polypeptides, e.g., a 21163 polypeptide having the amino acid sequence shown in SEQ ID NO:2; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, wherein the nucleic acid encodes a full length 21163 protein or an active fragment thereof.

In a related aspect, the invention further provides nucleic acid constructs which include a 21163 nucleic acid molecule described herein.

In a related aspect, the invention provides 21163 polypeptides or fragments operatively linked to non-21163 polypeptides to form fusion proteins.

In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically bind 21163 polypeptides.

In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 21163 polypeptides or nucleic acids.

In still another aspect, the invention provides a process for modulating 21163 polypeptide or nucleic acid expression or activity, e.g., using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 21163 polypeptides or nucleic acids, conditions such as anxiety, psychotic and depressive disorders, cognitive disorders, pain perception, neurodegenerative disorders, cancers, and infections such as HIV and malaria.

The invention also provides assays for determining the activity of or the presence or absence of 21163 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

In further aspect the invention provides assays for determining the presence or  
5 absence of a genetic alteration in a 21163 polypeptide or nucleic acid molecule,  
including for disease diagnosis.

#### BRIEF DESCRIPTION OF THE DRAWINGS

*Figures 1A, 1B, and 1C* depict a cDNA sequence (SEQ ID NO:1) and  
10 predicted amino acid sequence (SEQ ID NO:2) of human 21163. The methionine-  
initiated open reading frame of human 21163 (without the 5' and 3' untranslated  
regions) extends from nucleotide position 451 to position 2364 of SEQ ID NO:1, not  
including the terminal codon (coding sequence shown in SEQ ID NO:3).

*Figure 2* depicts a hydropathy plot of human 21163. Relative hydrophobic  
15 residues are shown above the dashed horizontal line, and relative hydrophilic  
residues are below the dashed horizontal line. The cysteine residues (cys) and N  
glycosylation site (Ngly) are indicated by short vertical lines just below the  
hydropathy trace. The numbers corresponding to the amino acid sequence (shown in  
SEQ ID NO:2) of human 21163 are indicated. Residues 463-479 resemble a  
20 transmembrane spanning segment. Polypeptides of the invention include fragments  
which include: all or a part of a hydrophobic sequence (a sequence above the dashed  
line); or all or part of a hydrophilic fragment (a sequence below the dashed line).  
Other fragments include a cysteine residue or an N-glycosylation site.

*Figure 3* depicts an alignment of portions of the prolyl oligopeptidase domain  
25 of human 21163 with consensus amino acid sequences derived from hidden Markov  
models. In the first alignment, the upper sequence is the consensus amino acid  
sequence for the prolyl oligopeptidase, N-terminal beta-propeller domain (PFAM  
Accession No. PF02897) and the lower amino acid sequences correspond to human  
21163. The upper sequence of the first alignment is set forth as SEQ ID NO:4 and the  
30 lower amino acid sequence corresponds to amino acid 1-342 of SEQ ID NO:2. In the  
second alignment, the upper sequence is the consensus amino acid sequences for the  
transcriptional regulatory protein, C terminal (PFAM Accession No. PF00486) and



the lower amino acid sequence corresponds to human 21163. The upper sequence of the second alignment is SEQ ID NO:5 and the lower amino acid sequence corresponds to amino acid 438-457 of SEQ ID NO:2. In the third alignment, the upper sequence is the consensus amino acid sequence for the Peptidase S9 family of prolyl oligopeptidases (PFAM Accession No. PF00326) and the lower amino acid sequences correspond to amino acids of human 21163. The upper sequence of the alignment is SEQ ID NO:6 and the lower amino acid sequence corresponds to amino acids 397 to 475 of SEQ ID NO:2.

*Figure 4* demonstrates the tumor suppressor induced down-regulation of 21163 expression. *Figure 4a* demonstrates that the induction of p53 expression by 4HT treatment resulted in a decrease expression of the 21163 mRNA over time. *Figure 4b* demonstrates that upon expression of p16, expression of the 2116 mRNA decreased. White boxes represent controls. Shaded boxes represent cells having elevated expression levels of p53 (4a) or p16 (4b).

*Figure 5* schematically illustrates the expression of 21163 in numerous tissue types.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention now will be described more fully hereinafter with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout.

Many modifications and other embodiments of the invention will come to mind to one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the invention is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to

be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

5    Human 21163

          The present invention provides the human 21163 sequence (Figure 1; SEQ ID NO:1), which is approximately 4959 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1914 nucleotides (nucleotides 451 to 2364 of SEQ ID NO:1; SEQ ID NO:3), not including the terminal  
10    codon. The coding sequence encodes a 638 amino acid protein (SEQ ID NO:2).

          Human 21163 (SEQ ID NO:2) contains the following regions or other structural features: a predicted prolyl oligopeptidase domain (PFAM Accession No. PF00326, Figure 3) located at about amino acid residues 397 to 475 of SEQ ID NO:2 and a prolyl oligopeptidase, N-terminal beta-propeller domain (PFAM Accession No.  
15    PF02897, Figure 3) located at about amino acids 1-342 of SEQ ID NO:2.

          The prolyl oligopeptidase family consists of a number of evolutionary related peptidases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent  
20    convergent evolution. A conserved serine residue has experimentally been shown (in *E. coli* protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine is part of the charge relaying catalytic triad (Ser, His, Asp). These residues are generally located within about 130 residues of the carboxyl-terminal extremity of these enzymes. These proteins belong to families Peptidase S9A, S9B, and S9C in the classification of peptidases (Rawlings & Barrett (1996),  
25    *Meth. Enzymol.* 244:19-61; [www.expasy.ch/cgi-bin/lists?peptidas.txt](http://www.expasy.ch/cgi-bin/lists?peptidas.txt)). Included in the prolyl oligopeptidase family are the prolyl endopeptidases (E.C. 3.4.21.26), the dipeptidyl peptidase IV's (EC. 3.4.14.5), and the acylamino-acid-releasing enzymes (acyl-peptide hydrolases; EC. 3.4.19.1).

          As used herein, the term "prolyl oligopeptidase" refers to a protein or  
30    polypeptide that is capable of catalyzing the cleavage of a polypeptide bond through hydrolysis; e.g., possessing polypeptide hydrolytic activity.

The 21163 protein contains a significant number of structural characteristics in common with members of the prolyl oligopeptidase family as described above. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence  
5      homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or  
10     mouse proteins. Members of a family can also have common functional characteristics.

As the biological functions of prolyl oligopeptidases include the regulation of protein turnover; the processing and degradation of bioactive peptides including hormones and neuropeptides; cellular growth and proliferation, differentiation, and  
15     apoptosis; T-cell activation; and cell-to-cell adhesion. In particular, prolyl oligopeptidases have been shown to have a role in anxiety, psychotic and depressive disorders, cognitive disorders, pain perception, neurodegenerative disorders, cancers, and infections such as HIV and malaria.

A 21163 polypeptide can include a "prolyl oligopeptidase" or regions  
20     homologous with the "Peptidase S9 family of prolyl oligopeptidases" or with the "propyl oligopeptidase, N-terminal beta-propeller" domain.

As used herein, the term "Peptidase S9 family of prolyl oligopeptidases" includes an amino acid sequence having a bit score for the alignment of the sequence to the Peptidase S9 family domain (HMM) of at least 8. Preferably, a Peptidase S9  
25     family domain has a bit score for the alignment of the sequence to the prolyl oligopeptidase domain (HMM) of at least 16 or greater. The Peptidase S9 family (HMM) has been assigned the PFAM Accession PF00326 (<http://pfam.wustl.edu/>). An alignment of the Peptidase S9 family domain of human 21163 (amino acids 397 to 475 of SEQ ID NO:2) with the consensus amino acid sequences derived from a  
30     hidden Markov model is depicted in Figure 3. The bit score for this alignment of 21163 is 45. As used herein, the term "propyl oligopeptidase N-terminal beta-propeller" domain includes an amino acid sequence having a bit score for the

alignment of the sequence to the prolyl oligopeptidase, N-terminal beta-propeller domain (HMM) of at least 8. Preferably, a N-terminal beta-propeller domain has a bit score for the alignment of the sequence to the N-terminal beta-propeller domain (HMM) of at least 16 or greater. The prolyl-oligopeptidase, N-terminal beta-  
5 propeller domain (HMM) has been assigned PFAM Accession No. PF02897 (<http://pfam.wust.edu/>). An alignment of the N-terminal beta-propeller domain of human 21163 (amino acids 1-342 of SEQ ID NO: 2) with the consensus amino acid sequences derived from a hidden Markov model is depicted in Figure 3. The bit score for this alignment of 21163 is -52.

10 In a preferred embodiment 21163 polypeptide or protein has regions with at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with the Peptidase S9 family of prolyl oligopeptidases (e.g., amino acid residues 397 to 475 of SEQ ID NO:2). In other embodiments the 21163 polypeptide or protein has regions with at least about 60%, 70%, 80%, 90%, 95%, 99% or 100% homology with the prolyl-  
15 oligopeptidase N-terminal beta-propeller domain (e.g., amino acid residues 1-342 of SEQ ID NO:2)

To identify the presence of a Peptidase S9 prolyl oligopeptidase region of homology or the presence of the prolyl-oligopeptidase N-terminal beta-propeller region of homology in a 21163 protein sequence, and make the determination that a  
20 polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters ([http://www.sanger.ac.uk/Software/Pfam/HMM\\_search](http://www.sanger.ac.uk/Software/Pfam/HMM_search)). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a  
25 family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth.*  
30 *Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

As the 21163 polypeptides of the invention may modulate 21163-mediated activities, they may be useful for developing novel diagnostic and therapeutic agents for 21163-mediated or related disorders, as described below.

As used herein, a "21163 activity", "biological activity of 21163" or "functional activity of 21163", refers to an activity exerted by a 21163 protein, polypeptide or nucleic acid molecule on e.g., a 21163-responsive cell or on a 21163 polypeptide substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 21163 activity is a direct activity, such as an association with a 21163 target molecule. A "target molecule" or "binding partner" or "ligand" or "substrate" is a molecule with which a 21163 protein binds or interacts in nature, e.g., a polypeptide that a 21163 protein cleaves. A 21163 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 21163 protein with a 21163 ligand. For example, the 21163 proteins of the present invention can have one or more of the following activities: 1) regulation of intracellular protein turnover; 2) degradation and processing of bioactive peptides including peptide hormones and neuropeptides; 3) proteolysis of beta-amyloid precursor protein; 4) T-cell activation; 5) cell-to-cell adhesion; 6) regulation of cellular growth, proliferation, differentiation ; 7) activation of signal transduction; and 8) cell matrix adherence.

Accordingly, 21163 protein may mediate various disorders, including anxiety, psychotic and depressive disorders, cognitive disorders, pain perception, neurodegenerative disorders, cancers, and infections such as HIV and malaria.

Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast, ovary and liver origin.

As used herein, the terms "cancer", "hyperproliferative" and "neoplastic" refer to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state. The term is meant to include all

types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. "Pathologic hyperproliferative" cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic

5 hyperproliferative cells include proliferation of cells associated with wound repair.

The terms "cancer" or "neoplasms" include malignancies of the various organ systems, such as affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors,  
10 non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus.

The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast  
15 carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which  
20 the tumor cells form recognizable glandular structures.

The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

The 21163 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of proliferative disorders. E.g., such disorders include  
25 hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia.  
30 Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L. (1991) *Crit. Rev. in*

*Oncol./Hematol. 11:267-97*); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms  
5 of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

The 21163 nucleic acids and proteins of the invention can also be used to treat  
10 and/or diagnose a variety of neurodegenerative diseases such as Alzheimer's disease, Lewy body disease, Parkinson's disease, Huntington's disease, Pick's disease, progressive supranuclear palsy, corticobasal degeneration, striatonigral degeneration, Shy-Drager syndrome, olivopontocerebellar atrophy, spinocerebellar ataxias, Friedreich ataxia, ataxia-telangiectasia, amyotrophic lateral sclerosis, bulbospinal  
15 atrophy (Kennedy syndrome), spinal muscular atrophy, Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, Canavan disease, and Leigh disease.

The 21163 sequence may also be used to treat and/or diagnose a variety of disorders that arise in the tissues in which it is expressed. Disorders involving the  
20 spleen include, but are not limited to, splenomegaly, including nonspecific acute splenitis, congestive splenomegaly, and splenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly include infections, such as nonspecific splenitis, infectious mononucleosis, tuberculosis, typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-  
25 azar, trypanosomiasis, schistosomiasis, leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic anemias, and thrombocytopenic purpura;  
30 immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick

disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by  
5 microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial  
10 (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis  
15 and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid,  
20 miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the colon include, but are not limited to, congenital  
25 anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery, infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory  
30 disorders, including parasites and protozoa, acquired immunodeficiency syndrome, transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such



as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease,  $\alpha_1$ -antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders involving the uterus and endometrium include, but are not limited to, endometrial histology in the menstrual cycle; functional endometrial disorders, such as anovulatory cycle, inadequate luteal phase, oral contraceptives and induced endometrial changes, and menopausal and postmenopausal changes; inflammations, such as chronic endometritis; adenomyosis; endometriosis; endometrial polyps; endometrial hyperplasia; malignant tumors, such as carcinoma of the endometrium;

mixed Müllerian and mesenchymal tumors, such as malignant mixed Müllerian tumors; tumors of the myometrium, including leiomyomas, leiomyosarcomas, and endometrial stromal tumors.

Disorders involving the brain include, but are not limited to, disorders  
 5 involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those  
 10 related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease,  
 15 including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and  
 20 neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral  
 25 neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic  
 30 encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including

Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive  
 supranuclear palsy, corticobasal degeneration, multiple system atrophy, including  
 striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy,  
 and Huntington disease; spinocerebellar degenerations, including spinocerebellar  
 5 ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases  
 affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron  
 disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy;  
 inborn errors of metabolism, such as leukodystrophies, including Krabbe disease,  
 metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher  
 10 disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh  
 disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic  
 diseases, including vitamin deficiencies such as thiamine (vitamin B<sub>1</sub>) deficiency and  
 vitamin B<sub>12</sub> deficiency, neurologic sequelae of metabolic disturbances, including  
 hypoglycemia, hyperglycemia, and hepatic encephalopathy, toxic disorders, including  
 15 carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate  
 and radiation-induced injury; tumors, such as gliomas, including astrocytoma,  
 including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic  
 astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma,  
 oligodendroglioma, and ependymoma and related paraventricular mass lesions,  
 20 neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other  
 parenchymal tumors, including primary brain lymphoma, germ cell tumors, and  
 pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic  
 syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma,  
 and malignant peripheral nerve sheath tumor (malignant schwannoma), and  
 25 neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including  
 Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous  
 sclerosis, and Von Hippel-Lindau disease.

Disorders involving T-cells include, but are not limited to, cell-mediated  
 hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated  
 30 cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus  
 erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies,  
 mixed connective tissue disease, and polyarteritis nodosa and other vasculitides;

- immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and
- 5 chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.
- 10 Diseases of the skin, include but are not limited to, disorders of pigmentation and melanocytes, including but not limited to, vitiligo, freckle, melasma, lentigo, nevocellular nevus, dysplastic nevi, and malignant melanoma; benign epithelial tumors, including but not limited to, seborrheic keratoses, acanthosis nigricans, fibroepithelial polyp, epithelial cyst, keratoacanthoma, and adnexal (appendage)
- 15 tumors; premalignant and malignant epidermal tumors, including but not limited to, actinic keratosis, squamous cell carcinoma, basal cell carcinoma, and merkel cell carcinoma; tumors of the dermis, including but not limited to, benign fibrous histiocytoma, dermatofibrosarcoma protuberans, xanthomas, and dermal vascular tumors; tumors of cellular immigrants to the skin, including but not limited to,
- 20 histiocytosis X, mycosis fungoides (cutaneous T-cell lymphoma), and mastocytosis; disorders of epidermal maturation, including but not limited to, ichthyosis; acute inflammatory dermatoses, including but not limited to, urticaria, acute eczematous dermatitis, and erythema multiforme; chronic inflammatory dermatoses, including but not limited to, psoriasis, lichen planus, and lupus erythematosus; blistering (bullous)
- 25 diseases, including but not limited to, pemphigus, bullous pemphigoid, dermatitis herpetiformis, and noninflammatory blistering diseases: epidermolysis bullosa and porphyria; disorders of epidermal appendages, including but not limited to, acne vulgaris; panniculitis, including but not limited to, erythema nodosum and erythema induratum; and infection and infestation, such as verrucae, molluscum contagiosum,
- 30 impetigo, superficial fungal infections, and arthropod bites, stings, and infestations.

In normal bone marrow, the myelocytic series (polymorphoneuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-

30%. Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes),

5 polymorphoneuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure

10 2-8) of *Immunology, Immunopathology and Immunity*, Fifth Edition, Sell *et al.* Simon and Schuster (1996), incorporated by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following: diseases involving hematopoietic stem cells; committed lymphoid progenitor cells; lymphoid cells

15 including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia, monocytic; [leukemias are encompassed with and without differentiation];

20 -chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute myelocytic leukemia, hematologic

25 malignancies of monocyte-macrophage lineage, such as juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis

30 localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell carcinoma; sarcoma, including kaposi's sarcoma; fibroadenoma and

phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium; demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyangitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboangitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

Disorders involving red cells include, but are not limited to, anemias, such as hemolytic anemias, including hereditary spherocytosis, hemolytic disease due to erythrocyte enzyme defects: glucose-6-phosphate dehydrogenase deficiency, sickle cell disease, thalassemia syndromes, paroxysmal nocturnal hemoglobinuria, immunohemolytic anemia, and hemolytic anemia resulting from trauma to red cells; and anemias of diminished erythropoiesis, including megaloblastic anemias, such as anemias of vitamin B12 deficiency: pernicious anemia, and anemia of folate deficiency, iron deficiency anemia, anemia of chronic disease, aplastic anemia, pure red cell aplasia, and other forms of marrow failure.

Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lymphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

Disorders involving B-cells include, but are not limited to precursor B-cell neoplasms, such as lymphoblastic leukemia/lymphoma. Peripheral B-cell neoplasms include, but are not limited to, chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, diffuse large B-cell lymphoma, Burkitt lymphoma, plasma cell neoplasms, multiple myeloma, and related entities, lymphoplasmacytic lymphoma (Waldenström macroglobulinemia), mantle cell lymphoma, marginal zone lymphoma (MALToma), and hairy cell leukemia.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include, but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble mediators, acute glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental



glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis,

5 glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including

10 but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to,

15 urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic

20 HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma

25 (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Disorders of the breast include, but are not limited to, disorders of development; inflammations, including but not limited to, acute mastitis, periductal mastitis, periductal

30 mastitis (recurrent subareolar abscess, squamous metaplasia of lactiferous ducts), mammary duct ectasia, fat necrosis, granulomatous mastitis, and pathologies associated with silicone breast implants; fibrocystic changes; proliferative breast disease including,

but not limited to, epithelial hyperplasia, sclerosing adenosis, and small duct papillomas; tumors including, but not limited to, stromal tumors such as fibroadenoma, phyllodes tumor, and sarcomas, and epithelial tumors such as large duct papilloma; carcinoma of the breast including in situ (noninvasive) carcinoma that includes ductal carcinoma in  
5 situ (including Paget's disease) and lobular carcinoma in situ, and invasive (infiltrating) carcinoma including, but not limited to, invasive ductal carcinoma, no special type, invasive lobular carcinoma, medullary carcinoma, colloid (mucinous) carcinoma, tubular carcinoma, and invasive papillary carcinoma, and miscellaneous malignant neoplasms.

Disorders in the male breast include, but are not limited to, gynecomastia and  
10 carcinoma.

Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to,  
15 gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumors of sex cord-gonadal stroma including, but not limited to, Leydig (interstitial) cell tumors and sertoli cell tumors  
20 (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

Disorders involving the thyroid include, but are not limited to, hyperthyroidism;  
25 hypothyroidism including, but not limited to, cretinism and myxedema; thyroiditis including, but not limited to, hashimoto thyroiditis, subacute (granulomatous) thyroiditis, and subacute lymphocytic (painless) thyroiditis; Graves disease; diffuse and multinodular goiter including, but not limited to, diffuse nontoxic (simple) goiter and multinodular goiter; neoplasms of the thyroid including, but not limited to, adenomas,  
30 other benign tumors, and carcinomas, which include, but are not limited to, papillary carcinoma, follicular carcinoma, medullary carcinoma, and anaplastic carcinoma; and congenital anomalies.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving the pancreas include those of the exocrine pancreas such as congenital anomalies, including but not limited to, ectopic pancreas; pancreatitis, including but not limited to, acute pancreatitis; cysts, including but not limited to, pseudocysts; tumors, including but not limited to, cystic tumors and carcinoma of the pancreas; and disorders of the endocrine pancreas such as, diabetes mellitus; islet cell tumors, including but not limited to, insulinomas, gastrinomas, and other rare islet cell tumors.

Disorders involving the small intestine include the malabsorption syndromes such as, celiac sprue, tropical sprue (postinfectious sprue), whipple disease, disaccharidase (lactase) deficiency, abetalipoproteinemia, and tumors of the small intestine including adenomas and adenocarcinoma.

Disorders related to reduced platelet number, thrombocytopenia, include idiopathic thrombocytopenic purpura, including acute idiopathic thrombocytopenic purpura, drug-induced thrombocytopenia, HIV-associated thrombocytopenia, and thrombotic microangiopathies: thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome.

Disorders involving precursor T-cell neoplasms include precursor T lymphoblastic leukemia/lymphoma. Disorders involving peripheral T-cell and natural killer cell neoplasms include T-cell chronic lymphocytic leukemia, large granular lymphocytic leukemia, mycosis fungoides and Sézary syndrome, peripheral T-cell lymphoma, unspecified, angioimmunoblastic T-cell lymphoma, angiocentric lymphoma (NK/T-cell lymphoma<sup>4a</sup>), intestinal T-cell lymphoma, adult T-cell leukemia/lymphoma, and anaplastic large cell lymphoma.

Disorders involving the ovary include, for example, polycystic ovarian disease, Stein-leventhal syndrome, Pseudomyxoma peritonei and stromal hyperthecosis; ovarian tumors such as, tumors of coelomic epithelium, serous tumors, mucinous tumors, endometrioid tumors, clear cell adenocarcinoma, cystadenofibroma, brenner tumor, surface epithelial tumors; germ cell tumors such as mature (benign) teratomas, monodermal teratomas, immature malignant teratomas, dysgerminoma, endodermal sinus tumor, choriocarcinoma; sex cord-stromal tumors

such as, granulosa-theca cell tumors, thecoma-fibromas, androblastomas, hill cell tumors, and gonadoblastoma; and metastatic tumors such as Krukenberg tumors.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and osteocytes. The disorders of the bone are complex because they may have an impact  
5 on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body. Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matix such as type 1 collagen disease, osteoporosis, Paget disease, rickets, osteomalacia, high-turnover osteodystrophy, low-  
10 turnover of aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocyto-  
15 tumor, and metastatic tumors.

The 21163 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 are collectively referred to as "polypeptides or proteins of the invention" or "21163 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the  
20 invention" or "21163 nucleic acids." 21163 molecules refer to 21163 nucleic acids, polypeptides, and antibodies.

As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid  
25 molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA,  
30 the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e.,

sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology* John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, or SEQ ID NO:3, corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 21163 protein, preferably a mammalian 21163 protein, and can further include non-coding regulatory sequences, and introns.

An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 21163 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-21163 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-21163 chemicals. When the 21163 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 21163 (e.g., the sequence of SEQ ID NO:1 or SEQ ID NO:3) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, in particular those present in the catalytic triad active site domain, are not predicted to be amenable to alteration.

A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains

(e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 21163 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 21163 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 21163 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

As used herein, a "biologically active portion" of a 21163 protein includes a fragment of a 21163 protein which participates in an interaction between a 21163 molecule and a non-21163 molecule. Biologically active portions of a 21163 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 21163 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 21163 proteins, and exhibit at least one activity of a 21163 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 21163 protein, e.g., prolyl oligopeptidase activity as described herein on pages 9-10. A biologically active portion of a 21163 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200, 300, 400, 500, 600 or more amino acids in length. Biologically active portions of a 21163 protein can be used as targets for developing agents which modulate a 21163 mediated activity, e.g., prolyl oligopeptidase activity.

Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a

reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 21163 amino acid sequence of  
5 SEQ ID NO:2 having 191 amino acid residues, at least 255, preferably at least 319, more preferably at least 383, even more preferably at least 447, and even more preferably at least 510, 574 or 638 amino acid residues are aligned. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same  
10 amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps,  
15 and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined  
20 using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences  
25 is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity  
30 or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.



The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (1989) *CABIOS* 4:11-17 which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

5       The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the  
10 NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 21163 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 21163 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST  
15 can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

"Misexpression or aberrant expression", as used herein, refers to a non-wild  
20 type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from  
25 wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular  
30 stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

"Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

A "purified preparation of cells", as used herein, refers to, in the case of plant  
5 or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

Various aspects of the invention are described in further detail below.

#### 10 Isolated Nucleic Acid Molecules

In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 21163 polypeptide described herein, e.g., a full length 21163 protein or a fragment thereof, e.g., a biologically active portion of 21163 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe,  
15 which can be used, e.g., to identify nucleic acid molecule encoding a polypeptide of the invention, 21163 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1 or a portion of any of these  
20 nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 21163 protein (i.e., "the coding region", from nucleotides 451-2364 of SEQ ID NO:1, not including the terminal codon), as well as 5' untranslated sequences (nucleotides 1-450 of SEQ ID NO:1). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (e.g.,  
25 nucleotides 451-2364 of SEQ ID NO:1, corresponding to SEQ ID NO:3) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the mature protein of SEQ ID NO:2.

In another embodiment, an isolated nucleic acid molecule of the invention  
30 includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is

sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3 thereby forming a stable duplex.

In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. In the case of an isolated nucleic acid molecule which is longer than or equivalent in length to the reference sequence, e.g., SEQ ID NO:1, or SEQ ID NO:3, the comparison is made with the full length of the reference sequence. Where the isolated nucleic acid molecule is shorter than the reference sequence, e.g., shorter than SEQ ID NO:1, or SEQ ID NO:3, the comparison is made to a segment of the reference sequence of the same length (excluding any loop required by the homology calculation).

#### 21163 Nucleic Acid Fragments

A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 21163 protein, e.g., an immunogenic or biologically active portion of a 21163 protein. A fragment can comprise all or a portion of the nucleotides from about nucleotide 1-50, 50-600, 200-600, 300-600, 600-800, 800-1100, 1100-1500, 1500-1800, 1800-2000, or 2000-2300 of SEQ ID NO:1, that encode a prolyl oligopeptidase domain of human 21163. The nucleotide sequence determined from the cloning of the 21163 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 21163 family members, or fragments thereof, as well as 21163 homologues, or fragments thereof, from other species.

In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly

fragments thereof which are at least 150 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not be construed as encompassing those fragments that may have been disclosed prior to the invention.

5           A nucleic acid fragment can include a sequence corresponding to a region or functional site described herein. A nucleic acid fragment can also include one or more regions or functional sites described herein. Thus, for example, a nucleic acid fragment can include a prolyl oligopeptidase domain. In a preferred embodiment the fragment is at least 50, 100, 200, 250, 300, 400, 600, 800, 1000, 1200, 1400, 1600, 1800 or more  
10   base pairs in length.

          21163 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60,  
15   65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1, SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3.

          In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length.  
20   It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

          A probe or primer can be derived from the sense or anti-sense strand of a  
25   nucleic acid which encodes a portion of a prolyl oligopeptidase consensus domain (e.g., about amino acid residues 397-475 of SEQ ID NO:2 or about amino acid residues 1-342 of SEQ ID NO:2).

          In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 21163 sequence,  
30   e.g., a region described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a

naturally occurring variant. E.g., primers suitable for amplifying all or a portion of any of a prolyl oligopeptidase consensus domain (e.g., about amino acid residues 397-475 of SEQ ID NO:2 or about amino acid residues 1-342 of SEQ ID NO:2).

5 A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

A nucleic acid fragment encoding a "biologically active portion of a 21163 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, which encodes a polypeptide having a 21163 biological activity (e.g., the biological activities of the 21163 proteins as described herein),  
10 expressing the encoded portion of the 21163 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 21163 protein. For example, a nucleic acid fragment encoding a biologically active portion of 21163 may include a polypeptide hydrolytic domain (e.g., about amino acid residues 397-530 of SEQ ID NO:2 or about amino acid residues 1-342 of SEQ ID NO:2). A nucleic acid  
15 fragment encoding a biologically active portion of a 21163 polypeptide, may comprise a nucleotide sequence that is 300, 400, 600, 800, 1000, 1200, 1400, 1600, 1800 or more nucleotides in length.

In preferred embodiments, nucleic acids include a nucleotide sequence that is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1600, or 1800  
20 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1 or SEQ ID NO:3.

#### 21163 Nucleic Acid Variants

The invention further encompasses nucleic acid molecules that differ from the  
25 nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 21163 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which  
30 differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that is shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences

should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Nucleic acids of the invention can be chosen for having codons, which are preferred, or non preferred, for a particular expression system. E.g., the nucleic acid  
5 can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non-  
10 naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the  
15 encoded product).

In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:1 or SEQ ID NO:3, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum  
20 homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about  
25 80-85%, and most typically at least about 90-95% or more identical to the amino acid sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be obtained as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:3 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of  
30 the 21163 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 21163 gene. Preferred variants include those that are correlated with prolyl oligopeptidase activity, e.g., variants that comprise nucleotide

sequences encoding polypeptides that share identity to the amino acid sequence shown in SEQ ID NO:2 or a fragment of this sequence retain prolyl oligopeptidase activity.

Allelic variants of 21163, e.g., human 21163, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 21163 protein within a population that maintain polypeptide hydrolytic activity as described herein. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 21163, e.g., human 21163, protein within a population that do not have the ability to catalyze the cleavage of peptide bonds. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

Moreover, nucleic acid molecules encoding other 21163 family members and, thus, which have a nucleotide sequence which differs from the 21163 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention.

#### Antisense Nucleic Acid Molecules, Ribozymes and Modified 21163 Nucleic Acid Molecules

In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 21163. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 21163 coding strand, or to only a portion thereof (e.g., the coding region of human 21163 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 21163 (e.g., the 5' and 3' untranslated regions).

An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 21163 mRNA, but more preferably is an oligonucleotide

which is antisense to only a portion of the coding or noncoding region of 21163 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 21163 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense  
5 oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be  
10 chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an  
15 expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically  
20 administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 21163 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically.  
25 For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient  
30 intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.



In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 21163-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 21163 cDNA disclosed herein (i.e., SEQ ID NO:1, or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 21163-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 21163 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

21163 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 21163 (e.g., the 21163 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 21163 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other,

eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent,  
5 radioactive, or colorimetric.

A 21163 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup  
10 B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low  
15 ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* *Proc. Natl. Acad. Sci.* 93:14670-675.

PNAs of 21163 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene  
20 agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 21163 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. (1996)  
25 *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad.*  
30 *Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci.* <sup>1</sup> *USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with

hybridization-triggered cleavage agents (See, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 21163 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 21163 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.* U.S. Patent No. 5,854,033; Nazarenko *et al.* U.S. Patent No. 5,866,336, and Livak *et al.* U.S. Patent No. 5,876,930.

#### Isolated 21163 Polypeptides

In another aspect, the invention features, an isolated 21163 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-21163 antibodies. 21163 protein can be isolated from cells or tissue sources using standard protein purification techniques. 21163 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same postranslational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of postranslational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

In a preferred embodiment, a 21163 polypeptide has one or more of the following characteristics:

- (i) it is capable of prolyl oligopeptidase activity;

- (ii) it has a molecular weight, e.g., a deduced molecular weight, amino acid composition or other physical characteristic of the polypeptide of SEQ ID NO:2;
- (iii) it has an overall sequence identity of at least 50%, preferably at least 60%, more preferably at least 70, 80, 90, 91%, 92%, 93%, 94%, 95%, 96%,  
5 97%, 98%, or 99%, with a polypeptide of SEQ ID NO:2;
- (iv) it has a prolyl oligopeptidase domain that preferably has an overall sequence identity of about 70%, 80%, 90%, or 95% or more with amino acid residues 397-475 of SEQ ID NO:2 or with amino acids 1-342 of SEQ ID NO:2;
- (v) it has at least 70%, preferably 80%, and most preferably 95% of  
10 the cysteines found amino acid sequence of the native protein.

In a preferred embodiment the 21163 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:2. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but less than 20%,  
15 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2. (If this comparison requires alignment, the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred  
20 embodiment the differences are not in the prolyl oligopeptidase domain. In another preferred embodiment one or more differences are in non-active site residues, e.g., prolyl oligopeptidase domain.

Other embodiments include a protein that contain one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for  
25 activity. Such 21163 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.

In one embodiment, a biologically active portion of a 21163 protein includes a prolyl oligopeptidase domain. In another embodiment, a biologically active portion of a 21163 protein includes a portion of the prolyl oligopeptidase hydrolytic domain  
30 that includes the catalytic triad. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for the functional activities of a native 21163 protein.

In a preferred embodiment, the 21163 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 21163 protein is substantially identical to SEQ ID NO:2. In yet another embodiment, the 21163 protein is substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2, as described in detail herein. Accordingly, in another embodiment, the 21163 protein is a protein which includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NO:2.

#### 10 21163 Chimeric or Fusion Proteins

In another aspect, the invention provides 21163 chimeric or fusion proteins. As used herein, a 21163 "chimeric protein" or "fusion protein" includes a 21163 polypeptide linked to a non-21163 polypeptide. A "non-21163 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 21163 protein, e.g., a protein which is different from the 21163 protein and which is derived from the same or a different organism. The 21163 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 21163 amino acid sequence. In a preferred embodiment, a 21163 fusion protein includes at least one biologically active portion of a 21163 protein. The non-21163 polypeptide can be fused to the N-terminus or C-terminus of the 21163 polypeptide.

The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-21163 fusion protein in which the 21163 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 21163. Alternatively, the fusion protein can be a 21163 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 21163 can be increased through use of a heterologous signal sequence.

Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

The 21163 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 21163 fusion

proteins can be used to affect the bioavailability of a 21163 substrate. 21163 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 21163 protein; (ii) misregulation of the 21163 gene; and (iii) aberrant post-translational modification of a 21163 protein. "Treatment" is herein defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A "therapeutic agent" includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

Moreover, the 21163-fusion proteins of the invention can be used as immunogens to produce anti-21163 antibodies in a subject, to purify 21163 ligands and in screening assays to identify molecules which inhibit the interaction of 21163 with a 21163 substrate.

Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 21163-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 21163 protein.

#### Variants of 21163 Proteins

In another aspect, the invention also features a variant of a 21163 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 21163 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 21163 protein. An agonist of the 21163 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 21163 protein. An antagonist of a 21163 protein can inhibit one or more of the activities of the naturally occurring form of the 21163 protein by, for example, competitively modulating a 21163-mediated activity of a 21163 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a

variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 21163 protein.

5 Variants of a 21163 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 21163 protein for agonist or antagonist activity.

Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 21163 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 21163 protein.

10 Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Recursive ensemble mutagenesis (REM), a new technique which  
15 enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 21163 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

Cell based assays can be exploited to analyze a variegated 21163 library. For  
20 example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 21163 in a substrate-dependent manner. The transfected cells are then contacted with 21163 and the effect of the expression of the mutant on signaling by the 21163 substrate can be detected, e.g., by measuring prolyl oligopeptidase hydrolytic activity as described herein. Plasmid DNA can then be  
25 recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 21163 substrate, and the individual clones further characterized.

In another aspect, the invention features a method of making a 21163 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 21163 polypeptide, e.g., a naturally  
30 occurring 21163 polypeptide. The method includes: altering the sequence of a 21163 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or

more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of making a fragment or analog of a 21163 polypeptide a biological activity of a naturally occurring 21163 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 21163 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

#### 10 Anti-21163 Antibodies

In another aspect, the invention provides an anti-21163 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin.

The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

20 A full-length 21163 protein or, antigenic peptide fragment of 21163 can be used as an immunogen or can be used to identify anti-21163 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 21163 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 21163. Preferably, 25 the antigenic peptide includes at least about 10, 15, 20, 30 or more amino acid residues.

Fragments of 21163 that include residues from about amino acid 518-550 of SEQ ID NO:2 can be used to make, e.g., used as immunogens, or characterize the specificity of an antibody or antibodies against what are believed to be hydrophillic regions of the 21163 protein. Similarly, a fragment of 21163 that includes residues from about amino acid 460-500 of SEQ ID NO:2 can be used to make an antibody



against what is believed to be a hydrophobic region of the 21163 protein; a fragment of 21163 that includes residues from about amino acid 435-475 of SEQ ID NO:2 can be used to make an antibody against the active site region of the 21163 protein.

Antibodies reactive with, or specific for, any of these regions, or other regions  
5 or domains described herein are provided.

In a preferred embodiment the antibody fails to bind an Fc receptor, e.g., it is a type which does not support Fc receptor binding or has been modified, e.g., by deletion or other mutation, such that it does not have a functional Fc receptor binding region.

10 Preferred epitopes encompassed by the antigenic peptide are regions of 21163 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 21163 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 21163 protein and  
15 are thus likely to constitute surface residues useful for targeting antibody production.

In a preferred embodiment the antibody binds an epitope on any domain or region on 21163 proteins described herein.

Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic  
20 treatment (and some diagnostic applications) of human patients.

The anti-21163 antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered (see, for example, Colcher, D. *et al.* (1999, Jun 30) *Ann. NY Acad. Sci.* 880:263-80; and Reiter, Y. (1996 Feb) *Clin. Cancer Res.* 2(2):245-52). The single chain antibody can be dimerized or multimerized to  
25 generate multivalent antibodies having specificities for different epitopes of the same target 21163 protein.

An anti-21163 antibody (e.g., monoclonal antibody) can be used to isolate 21163 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-21163 antibody can be used to detect 21163  
30 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-21163 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing

procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

15 Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

A vector can include a nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The

expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 21163 proteins, mutant forms of 21163 proteins, fusion proteins, and the like).

5           The recombinant expression vectors of the invention can be designed for expression of 21163 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in*  
10 *Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of  
15 either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity  
20 purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech  
25 Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be used in 21163 activity assays, (e.g., direct  
30 assays or competitive assays described in detail below), or to generate antibodies specific for 21163 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone

marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

To maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

The 21163 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166).

Developmentally-regulated promoters are also encompassed, for example, the murine

hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense  
5 orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of  
10 the regulation of gene expression using antisense genes see Weintraub, H. *et al.* (1986) Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1).

Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 21163 nucleic acid molecule within a recombinant  
15 expression vector or a 21163 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding  
20 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 21163 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or  
25 mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for  
30 introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium

phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

A host cell of the invention can be used to produce (i.e., express) a 21163 protein. Accordingly, the invention further provides methods for producing a 21163 protein using the host cells of the invention. In one embodiment, the method includes  
5 culturing the host cell of the invention (into which a recombinant expression vector encoding a 21163 protein has been introduced) in a suitable medium such that a 21163 protein is produced. In another embodiment, the method further includes isolating a 21163 protein from the medium or the host cell.

10 In another aspect, the invention features, a cell or purified preparation of cells which include a 21163 transgene, or which otherwise misexpress 21163. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 21163 transgene, e.g., a heterologous form of a 21163, e.g., a gene derived from  
15 humans (in the case of a non-human cell). The 21163 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous 21163, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders which are related to mutated or misexpressed 21163 alleles or for  
20 use in drug screening.

In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 21163 polypeptide.

Also provided are cells or a purified preparation thereof, e.g., human cells, in which an endogenous 21163 is under the control of a regulatory sequence that does  
25 not normally control the expression of the endogenous 21163 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 21163 gene. For example, an endogenous 21163 gene, e.g., a gene which  
30 is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell.

Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

## 5 Transgenic Animals

The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 21163 protein and for identifying and/or evaluating modulators of 21163 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 21163 gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 21163 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 21163 transgene in its genome and/or expression of 21163 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 21163 protein can further be bred to other transgenic animals carrying other transgenes.

21163 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under

the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

The invention also includes a population of cells from a transgenic animal, as  
5 discussed herein.

### Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening  
10 assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used, for example, to express a 21163 protein (e.g., via a recombinant expression vector in a host cell in  
15 gene therapy applications), to detect a 21163 mRNA (e.g., in a biological sample) or a genetic alteration in a 21163 gene, and to modulate 21163 activity, as described further below. The 21163 proteins can be used to treat disorders characterized by insufficient or excessive production of a 21163 substrate or production of 21163 inhibitors. In addition, the 21163 proteins can be used to screen for naturally  
20 occurring 21163 substrates, to screen for drugs or compounds which modulate 21163 activity, as well as to treat disorders characterized by insufficient or excessive production of 21163 protein or production of 21163 protein forms which have decreased, aberrant or unwanted activity compared to 21163 wild-type protein. Such disorders include those characterized by aberrant protein processing, protein  
25 degradation, cell growth or proliferation. Moreover, the anti-21163 antibodies of the invention can be used to detect and isolate 21163 proteins, regulate the bioavailability of 21163 proteins, and modulate 21163 activity.

A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 21163 polypeptide is provided. The method includes: contacting the  
30 compound with the subject 21163 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 21163 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in*



vivo, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 21163 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 21163 polypeptide. Screening methods are discussed in more detail below.

5

#### Screening Assays:

The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to  
10 21163 proteins, have a stimulatory or inhibitory effect on, for example, 21163 expression or 21163 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 21163 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 21163 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or  
15 to identify compounds that disrupt normal target gene interactions.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 21163 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the  
20 activity of a 21163 protein or polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries [libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive] (see, e.g., Zuckermann, R.N. *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are  
25 limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell that expresses a 21163 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 21163 activity is determined. Determining the ability of the test compound to modulate 21163 activity can be accomplished by monitoring, for example, prolyl oligopeptidase hydrolytic activity as described herein. The cell, for example, can be of mammalian origin, e.g., human. Cell homogenates, or fractions, preferably membrane containing fractions, can also be tested.

The ability of the test compound to modulate 21163 binding to a compound, e.g., a 21163 substrate, or to bind to 21163 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 21163 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 21163 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 21163 binding to a 21163 substrate in a complex. For example, compounds (e.g., 21163 substrates) can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radio emission or by scintillation counting.

Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (e.g., a 21163 substrate) to interact with 21163 with  
5 or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 21163 without the labeling of either the compound or the 21163. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies  
10 its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 21163.

In yet another embodiment, a cell-free assay is provided in which a 21163 protein or biologically active portion thereof is contacted with a test compound and  
15 the ability of the test compound to bind to the 21163 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 21163 proteins to be used in assays of the present invention include fragments which participate in interactions with non-21163 molecules, e.g., fragments with high surface probability scores.

20 Soluble and/or membrane-bound forms of isolated proteins (e.g., 21163 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide,  
25 Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl) dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl) dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

30 Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two

components to interact and bind, thus forming a complex that can be removed and/or detected.

In one embodiment, assays are performed where the ability of an agent to block prolyl oligopeptidase activity within a cell is evaluated.

5       The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.* U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.* U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able  
10       to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial  
15       relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

20       In another embodiment, determining the ability of the 21163 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without  
25       labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

30       In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene

product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize either 21163, an anti-21163 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 21163 protein, or interaction of a 21163 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/21163 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 21163 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads; complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 21163 binding or activity determined using standard techniques.

Other techniques for immobilizing either a 21163 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 21163 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any

complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the  
5 previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

In one embodiment, this assay is performed utilizing antibodies reactive with  
10 21163 protein or target molecules but which do not interfere with binding of the 21163 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 21163 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes  
15 using antibodies reactive with the 21163 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 21163 protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a  
20 number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P. (1993 Aug) *Trends Biochem Sci* 18(8):284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. *et al.* eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for  
25 example, Ausubel, F. *et al.* eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H. (1998 Winter) *J Mol. Recognit.* 11(1-6):141-8; Hage, D.S., and Tweed, S.A. (1997, Oct 10) *J. Chromatogr. B Biomed. Sci. Appl.* 699(1-2):499-525). Further, fluorescence energy transfer may also be  
30 conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

In a preferred embodiment, the assay includes contacting the 21163 protein or biologically active portion thereof with a known compound which binds 21163 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 21163 protein, wherein  
5 determining the ability of the test compound to interact with a 21163 protein includes determining the ability of the test compound to preferentially bind to 21163 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

The target gene products of the invention can, *in vivo*, interact with one or  
10 more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The  
15 preferred target genes/products for use in this embodiment are the 21163 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 21163 protein through modulation of the activity of a downstream effector of a 21163 target molecule. For example, the activity of the effector molecule on an appropriate target  
20 can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is  
25 prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction  
30 mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction,

but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex  
5 formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

These assays can be conducted in a heterogeneous or homogeneous format.  
10 Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test  
15 compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound  
20 to the reaction mixture after complexes have been formed. The various formats are briefly described below.

In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or  
25 indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is  
30 complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that



complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

In yet another aspect, the 21163 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 21163 ("21163-binding proteins" or "21163-bp") and are involved in 21163 activity. Such 21163-bps can be activators or inhibitors of signals by the

21163 proteins or 21163 targets as, for example, downstream elements of a 21163-mediated signaling pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 21163 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively, the 21163 protein can be fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 21163-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 21163 protein.

In another embodiment, modulators of 21163 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 21163 mRNA or protein evaluated relative to the level of expression of 21163 mRNA or protein in the absence of the candidate compound. When expression of 21163 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 21163 mRNA or protein expression. Alternatively, when expression of 21163 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 21163 mRNA or protein expression. The level of 21163 mRNA or protein expression can be determined by methods described herein for detecting 21163 mRNA or protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using

a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 21163 protein can be confirmed *in vivo*, e.g., in an animal.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 21163 modulating agent, an antisense 21163 nucleic acid molecule, a 21163-specific antibody, or a 21163-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

#### Detection Assays

Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome, e.g., to locate gene regions associated with genetic disease or to associate 21163 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

#### Chromosome Mapping

The 21163 nucleotide sequences or portions thereof can be used to map the location of the 21163 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 21163 sequences with genes associated with disease.

Briefly, 21163 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 21163 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 21163 sequences will yield an amplified fragment.

A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of

mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924).

Other mapping strategies e.g., in situ hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 21163 to a chromosomal location.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al. Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 21163 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any

unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR  
5 based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

#### Tissue Typing

10 21163 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA  
15 markers for RFLP (described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 21163 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These  
20 primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

Allelic variation occurs to some degree in the coding regions of these  
25 sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can provide  
30 positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding

sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 21163 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same  
5 reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

#### Use of Partial 21163 Sequences in Forensic Biology

10 DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of  
15 the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique  
20 to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 (e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for  
25 this use.

The 21163 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., a tissue containing prolyl oligopeptidase activity. This can be very useful in cases  
30 where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 21163 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., 21163 primers or probes can be used to screen tissue culture for contamination (i.e., screen for the presence of a mixture of different types of cells in a culture).

5

#### Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

10

Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 21163.

Such disorders include, e.g., a disorder associated with the misexpression of 21163, or a protein processing or protein degradation related disorder.

15

The method includes one or more of the following:

detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 21163 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region;

20

detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 21163 gene;

detecting, in a tissue of the subject, the misexpression of the 21163 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA;

25

detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 21163 polypeptide.

30

In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 21163 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence

which hybridizes to a sense or antisense sequence from SEQ ID NO:1 naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 21163 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 21163 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 21163.

Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

In preferred embodiments the method includes determining the structure of a 21163 gene, an abnormal structure being indicative of risk for the disorder.

In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 21163 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

#### Diagnostic and Prognostic Assays

The presence, level, or absence of 21163 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 21163 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 21163 protein such that the presence of 21163 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 21163 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 21163 genes; measuring the amount of protein encoded by the 21163 genes; or measuring the activity of the protein encoded by the 21163 genes.

The level of mRNA corresponding to the 21163 gene in a cell can be determined both by *in situ* and by *in vitro* formats.



The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule  
5 (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 21163 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500, 750, 1000 or more nucleotides in length and sufficient to specifically hybridize under stringent conditions to 21163 mRNA or  
10 genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an  
15 alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 21163 genes.

The level of mRNA in a sample that is encoded by one of 21163 can be  
20 evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis, 1987, U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.* U.S. Patent  
25 No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are  
30 from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents,

such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that  
5 can hybridize to mRNA that encodes the 21163 gene being analyzed.

In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 21163 mRNA, or genomic DNA, and comparing the presence of 21163 mRNA or genomic DNA in the control sample with the presence of 21163 mRNA or genomic DNA in the test sample.

10 A variety of methods can be used to determine the level of protein encoded by 21163. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment  
15 thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

20 The detection methods can be used to detect 21163 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 21163 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 21163 protein include  
25 introducing into a subject a labeled anti-21163 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 21163 protein, and comparing  
30 the presence of 21163 protein in the control sample with the presence of 21163 protein in the test sample.

The invention also includes kits for detecting the presence of 21163 in a biological sample. For example, the kit can include a compound or agent capable of detecting 21163 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further  
5 comprise instructions for using the kit to detect 21163 protein or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

10 For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein-  
15 stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with  
20 instructions for interpreting the results of the assays performed using the kit.

The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 21163 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as altered protein  
25 degradation, protein processing, cell growth or proliferation.

In one embodiment, a disease or disorder associated with aberrant or unwanted 21163 expression or activity is identified. A test sample is obtained from a subject and 21163 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 21163 protein or nucleic acid is  
30 diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 21163 expression or activity. As used herein, a

"test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 21163 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for altered protein degradation, protein processing related disorder, or a cell growth or cell proliferation disorder.

The methods of the invention can also be used to detect genetic alterations in a 21163 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 21163 protein activity or nucleic acid expression, such as an altered protein degradation or protein processing related disorder, or a cell growth or cell proliferation disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 21163-protein, or the misexpression of the 21163 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 21163 gene; 2) an addition of one or more nucleotides to a 21163 gene; 3) a substitution of one or more nucleotides of a 21163 gene; 4) a chromosomal rearrangement of a 21163 gene; 5) an alteration in the level of a messenger RNA transcript of a 21163 gene; 6) aberrant modification of a 21163 gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 21163 gene; 8) a non-wild type level of a 21163-protein; 9) allelic loss of a 21163 gene; and 10) inappropriate post-translational modification of a 21163-protein.

An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 21163-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both)

from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 21163 gene under conditions such that hybridization and amplification of the 21163-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

In another embodiment, mutations in a 21163 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in 21163 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two-dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in 21163 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long

stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller,  
5 specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 21163 gene and detect mutations by  
10 comparing the sequence of the sample 21163 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.* (1995) *Biotechniques* 19:448-453), including sequencing by mass spectrometry.

Other methods for detecting mutations in the 21163 gene include methods in  
15 which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242-1246; Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4397-4401; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

In still another embodiment, the mismatch cleavage reaction employs one or  
20 more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 21163 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994)  
25 *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 21163 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc. Natl. Acad. Sci.*  
30 *USA*: 86:2766-2770, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 21163 nucleic acids will be denatured and allowed to renature. The

secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in  
5 which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in  
10 polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495-498). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature  
15 gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or  
20 selective primer extension (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6230).

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of  
25 interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the  
30 mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1-7). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci.*

USA 88:189-193). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing  
5 pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 21163 gene.

#### 10 Use of 21163 Molecules as Surrogate Markers

The 21163 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the  
15 presence, absence and/or quantity of the 21163 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 21163 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which  
20 correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular  
25 use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using  
30 HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of



surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The 21163 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 21163 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-21163 antibodies may be employed in an immune-based detection system for a 21163 protein marker, or 21163-specific radiolabeled probes may be used to detect a 21163 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90:229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S16-S20.

The 21163 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical

marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA or protein (e.g., 21163 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 21163 DNA may correlate 21163 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

#### Pharmaceutical Compositions

The nucleic acid and polypeptides, fragments thereof, as well as anti-21163 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic

solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be  
5 adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the  
10 extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of  
15 manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating  
20 such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as  
25 mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination  
30 of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other

ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

5           Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant  
10           materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant  
15           such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

20           Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished  
25           through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

30           The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled

release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and  
5 Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No.  
10 4,522,811.

It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated  
15 to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the  
20 dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in  
25 order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with  
30 little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be

estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids),

amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not

limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-  
5 dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological  
10 response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth  
15 factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

20 Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent No.  
25 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g.,  
30 retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.



The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment:

5           The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 21163 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of  
10   pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype" or "drug response  
15   genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 21163 molecules of the present invention or 21163 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the  
20   treatment and to avoid treatment of patients who will experience toxic drug-related side-effects.

          In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 21163 expression or activity, by administering to the subject a 21163 or an agent which modulates 21163  
25   expression or at least one 21163 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 21163 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 21163 aberrance, such that a disease  
30   or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 21163 aberrance, for example, a 21163, 21163 agonist or 21163 antagonist

agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

It is possible that some 21163 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting  
5 abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

As discussed, successful treatment of 21163 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that  
10 proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 21163 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')<sub>2</sub> and  
15 FAb expression library fragments, scFV molecules, and epitope-binding fragments thereof).

Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still  
20 further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal  
25 target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable  
30 to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 21163 expression is through the use of aptamer molecules specific for 21163 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, *et al.* (1997) *Curr. Opin. Chem. Biol.* 1(1):5-9; and Patel, D.J. (1997 Jun) *Curr. Opin. Chem. Biol.* 1(1):32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 21163 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of 21163 disorders. For a description of antibodies, see the Antibody section above.

In circumstances wherein injection of an animal or a human subject with a 21163 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 21163 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D. (1999) *Ann. Med.* 31(1):66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A. (1998) *Cancer Treat. Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 21163 protein. Vaccines directed to a disease characterized by 21163 expression may also be generated in this fashion.

In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered,

for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

5       The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 21163 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

10       Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Compounds that exhibit large therapeutic indices are preferred. While 15       compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

      The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds 20       lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to 25       achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

30       Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have

been created through molecular imprinting techniques. The compound which is able to modulate 21163 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. *et al.* (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J. (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. *et al.* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 21163 can be readily monitored and used in calculations of  $IC_{50}$ .

Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual  $IC_{50}$ . A rudimentary example of such a "biosensor" is discussed in Kriz, D. *et al.* (1995) *Analytical Chemistry* 67:2142-2144.

Another aspect of the invention pertains to methods of modulating 21163 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 21163 or agent that modulates one or more of the activities of 21163 protein activity associated with the cell. An agent that modulates 21163 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 21163 protein (e.g., a 21163 substrate or receptor), a 21163 antibody, a 21163 agonist or antagonist, a peptidomimetic of a 21163 agonist or antagonist, or other small molecule.

In one embodiment, the agent stimulates one or 21163 activities. Examples of such stimulatory agents include active 21163 protein and a nucleic acid molecule

encoding 21163. In another embodiment, the agent inhibits one or more 21163 activities. Examples of such inhibitory agents include antisense 21163 nucleic acid molecules, anti-21163 antibodies, and 21163 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, 5 *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 21163 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination 10 of agents that modulates (e.g., upregulates or downregulates) 21163 expression or activity. In another embodiment, the method involves administering a 21163 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 21163 expression or activity.

Stimulation of 21163 activity is desirable in situations in which 21163 is 15 abnormally downregulated and/or in which increased 21163 activity is likely to have a beneficial effect. For example, stimulation of 21163 activity is desirable in situations in which a 21163 is downregulated and/or in which increased 21163 activity is likely to have a beneficial effect. Likewise, inhibition of 21163 activity is desirable in situations in which 21163 is abnormally upregulated and/or in which 20 decreased 21163 activity is likely to have a beneficial effect.

The 21163 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of anxiety, psychotic and depressive disorders, cognitive disorders, pain perception, neurodegenerative disorders, cancers, and infections such as HIV and malaria.

25 Similarly, aberrant expression and/or activity of 21163 molecules may mediate disorders associated with, for example, the regulation of protein turnover; the processing and degradation of bioactive peptides including hormones and neuropeptides; cellular growth and proliferation, differentiation, and apoptosis; T-cell activation; and cell-to-cell adhesion.

### Pharmacogenomics

The 21163 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 21163 activity (e.g., 21163 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 21163 associated disorders (e.g., cellular growth related disorders) associated with aberrant or unwanted 21163 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 21163 molecule or 21163 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 21163 molecule or 21163 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related

markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. 5 Alternatively, such a high-resolution map can be generated from a combination of some ten million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 10 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into 15 account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 21163 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it 20 can be determined if having one version of the gene versus another is associated with a particular drug response.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 21163 molecule or 21163 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been 25 turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied 30 to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 21163



molecule or 21163 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 21163 genes of the present invention, wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 21163 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., cancer cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 21163 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 21163 gene expression, protein levels, or upregulate 21163 activity, can be monitored in clinical trials of subjects exhibiting decreased 21163 gene expression, protein levels, or downregulated 21163 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 21163 gene expression, protein levels, or downregulate 21163 activity, can be monitored in clinical trials of subjects exhibiting increased 21163 gene expression, protein levels, or upregulated 21163 activity. In such clinical trials, the expression or activity of a 21163 gene, and preferably, other genes that have been implicated in, for example, a 21163-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

## 25 Other Embodiments

In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a 21163, preferably purified, nucleic acid, preferably purified, polypeptide, preferably purified,

or antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the 21163 nucleic acid, polypeptide, or antibody.

5           The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

          The method can include contacting the 21163 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can  
10 be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or  
15 stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

          The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 21163. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a  
20 disease or disorder. 21163 is associated with prolyl oligopeptidase activity, thus it is useful for disorders associated with the regulation of protein turnover; the processing and degradation of bioactive peptides including hormones and neuropeptides; cellular growth, proliferation, differentiation, and apoptosis; T-cell activation; and cell-to-cell adhesion.

25           The method can be used to detect SNPs, as described above.

          In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of  
30 the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express or misexpress 21163 or from a cell or subject in which a 21163 mediated response has been elicited, e.g., by contact of the cell with 21163

nucleic acid or protein, or administration to the cell or subject 21163 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 21163 nucleic acid, polypeptide, or antibody); providing a two dimensional array  
5 having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 21163 (or does not express as highly as in the case of the 21163 positive plurality of capture probes) or from a cell or subject which  
10 in which a 21163 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 21163 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the  
15 plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

In another aspect, the invention features, a method of analyzing 21163, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 21163 nucleic acid or amino acid  
20 sequence; comparing the 21163 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 21163.

Preferred databases include GenBank™. The method can include evaluating the sequence identity between a 21163 sequence and a database sequence. The  
25 method can be performed by accessing the database at a second site, e.g., over the internet.

In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 21163. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an  
30 interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with

different labels, such that an oligonucleotide which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide which hybridizes to a second allele.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

### EXAMPLES

#### 10 Example 1: Identification and Characterization of Human 21163 cDNAs

The human 21163 sequence (Figure 1A-B; SEQ ID NO:1), which is approximately 4959 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1914 nucleotides (nucleotides 451-2364 of SEQ ID NO:1; SEQ ID NO:3). The coding sequence  
15 encodes a 638 amino acid protein (SEQ ID NO:2).

An analysis of the full-length 21163 polypeptide by MEMSAT predicts a transmembrane segment from amino acids (aa) 463 to 479. Prosite program analysis was used to predict various sites within the 21163 protein. An N-glycosylation site was predicted at about aa 112-115 and about aa 182-185. Cyclic AMP and cyclic  
20 GMP- dependant protein kinase phosphorylation sites were predicted at about aa 255-258 and about aa 608-611. Protein kinase C phosphorylation sites were predicted at about aa 157-159, 253-255, 291-293, 294-296, 330-332, 351-353, 437-439, 568-570, and 606-608. Casein kinase II phosphorylation sites were predicted at about aa 9-12, 43-46, 114-117, 145-148, 157-160, 184-187, 195-198, 238-241, 308-311, 342-345,  
25 375-378, 482-485, 502-505, 511-514, 522-525, 551-554, 566-569, and 627-630. Tyrosine kinase phosphorylation sites at about aa 73-80 and 150-156. N-myristoylation sites were predicted at aa 423-428, 455-460, 472-477, 519-524, and 622-627.

In addition, the 21163 sequence share sequence identity with the prolyl  
30 oligopeptidase, N-terminal beta-propeller domain and the Peptidase S9 family of prolyl oligopeptidases as predicted by HMMer, Version 2.

For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and <http://www.psc.edu/general/software/packages/pfam/pfam.html>. The alignments are shown in Figure 3.

5

#### Example 2: Tissue Distribution of 21163 mRNA

Northern blot hybridizations with various RNA samples can be performed under standard conditions and washed under stringent conditions, i.e., 0.2 X SSC at 65°C. A DNA probe corresponding to all or a portion of the 21163 cDNA (SEQ ID NO:1) can be used. The DNA is radioactively labeled with <sup>32</sup>P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

15

#### Example 3: Recombinant Expression of 21163 in Bacterial Cells

In this example, 21163 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 21163 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-21163 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

20

25

#### Example 4: Expression of Recombinant 21163 Protein in COS Cells

To express the 21163 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV

30

promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 21163 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the  
5 recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 21163 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 21163 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other  
10 restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 21163 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 21163 gene  
15 is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 $\alpha$ , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

20 COS cells are subsequently transfected with the 21163-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring*  
25 *Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 21163 polypeptide is detected by radiolabelling ( $^{35}\text{S}$ -methionine or  $^{35}\text{S}$ -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA  
30 specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with  $^{35}\text{S}$ -methionine (or  $^{35}\text{S}$ -cysteine). The culture media are then collected and the cells are

lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

5           Alternatively, DNA containing the 21163 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 21163 polypeptide is detected by radiolabelling and immunoprecipitation using a 21163 specific monoclonal antibody.

10

#### Example 5. Expression Patterns of the 21163 Sequence

The p53 tumor suppressor gene has been the subject of intense study for a number of years. In addition to its well defined role in transcriptional activation, p53 can also act to suppress the transcription of a number of genes involved in cellular proliferation. See, for example, Badie *et al.* (2000) *Mol Cell Biol.* 20:2358 and Zhao *et al.* (2000) *Genes Dev* 14:981 both of which are herein incorporated by reference.

To study the effects of p53 on the expression level of the 21163 sequence, a p53/estrogen receptor fusion protein (p53ER) was introduced into a lung tumor cell line that is null for the p53 protein. The p53 activity of this fusion protein is induced by addition of the estrogen analogue tamoxifen (4HT) to the cell culture medium.

Sequences downregulated upon expression of p53 may contribute to the process of cell transformation.

Figure 4 illustrates the effects of p53 activation on the expression levels of the 21163 sequence in a p53 null lung tumor cell line. Specifically, expression of 21163 was repressed upon activation of an engineered p53/estrogen-receptor fusion protein in H125 cells (Figure 4a). Consistent with this result, a correlation between expression of the p16 tumor suppressor and reduced levels of 21163 mRNA was also demonstrated (Figure 4b). The p16 tumor suppressor, which like p53, can act to induce a G1 cell cycle arrest in non-dividing cells.

30           The expression levels of the 21163 sequences across a range of normal and cancerous tissues were also determined. Figure 5 schematically illustrates the expression in numerous tissue types. The expression data shown in Figure 5 is

summarized in Table 1. While expression of the 21163 sequence was seen in a range of normal tissues, the highest expression occurred in the central nervous system and the skeletal muscle.



Table I. Phase I Taqman Ct and Expression Values for 21163 Expression

Tissue Type	Mean	S 2 Mean	ΔΔ Ct	Expression
Artery normal	31.48	23.95	7.54	5.3732
Aorta diseased	33.05	22.85	10.2	0.8531
Vein normal	29.27	20.84	8.43	2.9095
Coronary SMC	32.01	23.36	8.65	2.4894
HUVEC	29.16	21.86	7.3	6.3238
Hemangioma	31.09	20.16	10.93	0.5143
Heart normal	28.27	21.05	7.23	6.6612
Heart CHF	26.14	20.25	5.88	16.9802
Kidney	27.48	20.79	6.7	9.6183
Skeletal Muscle	29.52	24.76	4.76	36.906
Adipose normal	34.1	22.03	12.07	0.2318
Pancreas	31	23.13	7.88	4.2598
primary osteoblasts	34.41	22.25	12.16	0.2185
Osteoclasts (diff)	39.09	19.19	19.9	0
Skin normal	34.15	23.07	11.09	0.4603
Spinal cord normal	31.52	22.34	9.19	1.7181
Brain Cortex normal	25.41	23.36	2.06	240.6486
Brain Hypothalamus normal	25.85	23.59	2.27	208.0497
Nerve	32.5	23.09	9.41	1.47
*DRG (Dorsal Root Ganglion)	27.52	23.37	4.14	56.5237
Breast normal	31.66	22.41	9.25	1.6424
Breast tumor	33.44	22.22	11.23	0.4178
Ovary normal	27.52	21.83	5.7	19.3034
Ovary Tumor	28.03	21.61	6.42	11.6785
Prostate Normal	29.14	21.36	7.78	4.5497
Prostate Tumor	30.69	21.84	8.85	2.1671
Salivary glands	35.84	21.18	14.66	0
**Colon normal	32.52	19.86	12.66	0.154
Colon Tumor	30.96	22.82	8.13	3.5573
Lung normal	32.22	19.54	12.68	0.1529
Lung tumor	27.7	21.62	6.08	14.8335
Lung COPD	31.57	19.84	11.73	0.2944
Colon IBD	34	19.12	14.88	0.0332
Liver normal	33.78	21.57	12.22	0.2103
Liver fibrosis	31.45	23.11	8.34	3.0861
Spleen normal	33.08	21.22	11.86	0.269
Tonsil normal	31.22	18.68	12.55	0.1673
Lymph node normal	32.07	20.68	11.4	0.3713
Small intestine normal	32.95	21.76	11.19	0.4295
Macrophages	40	18.55	21.45	0
Synovium	33.9	21.14	12.77	0.1437
BM-MNC	38.36	20.19	18.17	0
Activated PBMC	34.84	19.12	15.72	0.0186
Neutrophils	39.07	20.36	18.71	0
Megakaryocytes	32.83	19.75	13.07	0.1163
Erythroid	32.69	23.14	9.55	1.3294
positive control	26.94	21.48	5.46	22.7972

The expression levels of the 21163 sequences was further analyzed in numerous other tumorous and normal (N) tissue types. The expression data from these experiments is summarized in Table 2. Increased expression of the 21163 in tumors of the breast, lung, and colon compared to their normal counterparts was seen.

Table 2. Expression Levels of 21163 in Oncology Phase II Panel

Tissue Type	21163.1 Mean	B 2 Mean	ΔΔ Ct	Expression
PIT 400 Breast N	31.08	21.02	10.06	0.94
ONC 038 Breast N	32.2	24.15	8.05	3.76
CHT 1228 Breast N	27.86	22.04	5.82	17.76
NDR 005 Breast Tum:IDC-MD/PD	27.86	22.2	5.67	19.71
CHT 2002 Breast T: IDC	27.77	20.91	6.87	8.55
CHT 564 Breast Tum:IDC-PD	27.66	24.56	3.11	115.82
CHT 562 Breast T: IDC	27.3	20	7.29	6.39
NDR 138 Breast T ILC (LG)	27.13	21.54	5.59	20.76
CHT 1841 Lymph node (Breast met)	28.92	22.63	6.29	12.74
PIT 58 Lung (Breast met)	31.66	23.68	7.97	3.97
CHT 620 Ovary N	27.77	20.77	7	7.81
CHT 619 Ovary N	25.97	21.15	4.83	35.28
CLN 012 Ovary T: PD-PS	27.11	22.08	5.03	30.61
CHT 2432 Ovary T:MD-PS	26.11	20.47	5.63	20.12
CLN 17 Ovary T: PD-PS	26.43	21.29	5.14	28.26
CHT 2434 Ovary T:PD-AC	26.11	20.26	5.84	17.40
CLN 08 Ovary T: MD/PD-PS	27.55	19.43	8.12	3.59
PIT 298 Lung N	28.95	20.66	8.29	3.21
PIT 270 Lung N	29.25	19.61	9.65	1.24
CLN 930 Lung N	31.73	22.19	9.54	1.34
MPI-215 Lung T-SmC	26.4	21.44	4.96	32.13
CHT 793 Lung T: MD-SCC	28.6	20.11	8.49	2.77
CHT 832 Lung T: PD-NSCLC	28.81	21.04	7.77	4.58
CHT 211 Lung T: WD-AC	29.19	21.17	8.02	3.85
CHT 1371 Lung T: MD-AC	31.42	20.97	10.46	0.71
CHT 331 Lung T: MD-AC	27.98	22.92	5.05	30.08
NDR 104 Colon N	26.97	19.7	7.27	6.48
CHT 1685 Colon N	28.27	20.66	7.61	5.12
CHT 371 Colon N	30.41	21.38	9.04	1.91
CHT 382 Colon T: MD	31.11	22.73	8.38	3.00
CHT 528 Colon T: MD	25.93	19.57	6.37	12.13
CLN 609 Colon T	28.16	20.46	7.71	4.79
NDR 210 Colon T: MD-PD	31.24	24.51	6.73	9.42
CHT 340 Colon-Liver Met	27.75	22.34	5.41	23.60
CHT 1637 Colon-Liver Met	28.43	21.57	6.86	8.64
PIT 260 Liver N (female)	29.36	20.09	9.27	1.62
CHT 1653 Cervix Squamous CC	27.93	22.36	5.57	21.05
CHT 569 Cervix Squamous CC	31.36	20.36	11	0.49
A24 HMVEC-Arr	27.27	20.48	6.79	9.04
C48 HMVEC-Prol	26.43	20.61	5.82	17.76
Pooled Hemangiomas	30.52	20.9	9.62	1.27
HCT116N22 Normoxic	26.52	23.05	3.48	89.93
HCT116H22 Hypoxic	28.82	23.51	5.31	25.21

The expression of the 21163 sequence in both normal and tumor epithelium of the lung and ovary was also examined. As summarized in Table 3, largely tumor specific expression in ovarian epithelium was observed.

5 Table 3. Summary of 21163 ISH Results

Spectrum	Tissue	Diagnosis	Results
<b>LUNG: 2/3 normals; 8/9 tumors</b>			
PIT 270	Lung	Normal	(+/-)
CHT 688	Lung	Normal	-
CHT 1240	Lung	Normal	(+/-)
CHT 446	Lung	Adenocarcinoma	(+/-)
CHT 211	Lung	Adenocarcinoma	(+/-)
CHT 331	Lung	Adenocarcinoma	-
CHT 1816	Lung	SCC	(+/-)
CHT 1814	Lung	SCC	-
CHT 344	Lung	SCC	(-/++)
MPI 226	Lung	Small Cell Lung CA	-
CHT 1812	Lung	PD-NSCLC	(+/-)
CHT 832	Lung	PD-NSCLC	(+)
<b>OVARY: 1/2 normals; 5/6 tumors</b>			
CLN 571	Ovary	Normal	-
CHT 619	Ovary	Normal	(+/-)
CLN 572	Ovary	Normal	-
CHT 2431	Ovary	Adenocarcinoma	(+/++)
CHT 2436	Ovary	Adenocarcinoma	(+/++)
CHT 2430	Ovary	Adenocarcinoma	-
CHT 2429	Ovary	Adenocarcinoma	(+/++)
CHT 2432	Ovary	Adenocarcinoma	+
CLN 6	Ovary	Adenocarcinoma	(+/++)

And finally, the expression of the 21163 sequences in lung xenograft cell lines was also examined. The results are summarized below in Table 4.

Table 4. 21163 Expression in Lung Xenograft

Tissue Type	21163 Mean	$\beta$ 2 Mean	$\Delta\Delta$ Ct	Expression
NHBE/ MPI003142	25.7	21.68	4.03	61.43
A549 / MPI003143	25.63	22.34	3.29	102.59
H460 / MPI003144	24.26	20.45	3.81	71.55
H23/ MPI003145	24.34	22.21	2.13	227.67
H522/ MPI003146	23.4	20.13	3.27	104.02
H125/ MPI003147	24.25	20.72	3.54	86.27
H520/ MPI003148	25.3	21.17	4.13	57.11
H69/ MPI003149	24.84	21.48	3.37	97.06
H345/ MPI003150	24.38	21.77	2.61	163.80
H460 INCX 24HR/ MPI003151	25.12	20.52	4.59	41.38
H460 p16 24HR/ MPI003152	28.17	22.62	5.55	21.34
H460 INCX STABLE PLAS/ MPI003155	24.37	20.32	4.05	60.37
H460 p16 STABLE PLAS/ MPI003156	24.75	21.2	3.56	85.08
H460 INCX STABLE AGAR/ MPI003158	24.36	19.06	5.3	25.30
H460 p16 STABLE AGAR/ MPI003159	25.07	20.25	4.82	35.53
H125 INCX 96HR/ MPI003160	26.31	22.39	3.92	66.06
H125 p53 96HR/ MPI003161	26.53	22.57	3.96	64.48
H125 p53ER 0hr/MPC001080	23.66	19.23	4.43	46.39
H125 p53ER 8hr/MPC001081	23.39	18.94	4.45	45.91
H125 p53ER 24hr/MPC001082	23.73	19.4	4.34	49.55
H125 p53ER 48hr/MPC001083	23.39	19.06	4.33	49.72
H125 p53ER 96hr/MPC001084	23.52	18.8	4.72	37.81
H125 VC 0hr/ MPC001085	23.49	19.59	3.9	66.99
H125 VC 8hr/ MPC001086	23.93	19.63	4.3	50.77
H125 VC 24hr/ MPC001087	24.02	19.75	4.27	51.83
H125 VC 48hr/ MPC001088	23.69	19.6	4.09	58.72
H125 VC 96hr/ MPC001089	24.11	19.95	4.17	55.55
H345 SPA 0hr/ MPC001117	23.13	21.06	2.06	238.99
H345 SPA 20min/ MPC001118	23.11	20.93	2.19	219.15
H345 SPA 1hr/ MPC001119	23.32	21	2.32	200.27

H345 SPA 3hr/ MPC001120	23.66	21.39	2.27	206.61
H345 SPA 9hr/ MPC001121	23.42	21.54	1.88	271.68
H345 SPA 21hr/ MPC001122	24.43	22.86	1.56	337.98
H69 SPA 0hr/ MPC000504	26.29	22.43	3.86	68.87
H69 SPA 1hr/ MPC000505	23.93	20.52	3.42	93.75
H69 SPA 6hr/ MPC000506	25.23	21.22	4.01	61.85
H69 SPA 12hr/ MPC000507	25.52	21.59	3.92	65.84
H69 SPA 24hr/ MPC000508	26.11	21.9	4.22	53.66
H69 SPA 36hr/ MPC000509	24.38	21.29	3.09	117.85

In summary, 21163 mRNA expression is increased in tumor samples and reduced upon activation of p53 and p16 in lung tumor cell lines that normally lack expression of these tumor suppressors. A number of genes that are regulated in this fashion have been shown to be critical for cell proliferation and survival (ex. cyclin A, thymidine kinase, 14-3-3), and 21163 my fall into this class of genes. Therefore small molecule therapeutics designed to inhibit the enzymatic activity of 21163 would be expected to reduce proliferation and survival of tumor cells. Additional support for a role of 21163 in tumorigenesis comes from literature reports showing increased prolyl oligopeptidase activity in lung tumors<sup>1</sup> and proliferating cells. See, for example, Sedo *et al.* (1991) *J Cancer Res Clin Oncol.* 117(3):249-253, Ohtuski *et al.* (1997) *J Biochem (Tokyo)*. 121(6):1176-1181, and Matsubara *et al.* (1998) *Eur J Biochem.* 252(1):178-183, all of which are herein incorporated by reference.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

## THAT WHICH IS CLAIMED:

1. An isolated nucleic acid molecule selected from the group consisting of:
  - 5 a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3;
  - b) a nucleic acid molecule comprising a fragment of at least 30 nucleotides of the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3;
  - 10 c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2;
  - d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2; and,
  - 15 e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or a complement thereof, under stringent conditions.
- 20 2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:
  - a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3; and
  - b) a nucleic acid molecule which encodes a polypeptide
  - 25 comprising the amino acid sequence of SEQ ID NO:2.
3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.
- 30 4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.



5. A host cell which contains the nucleic acid molecule of claim 1.
6. The host cell of claim 5 which is a mammalian host cell.
- 5 7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.
8. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide which is encoded by a nucleic acid molecule  
10 comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a complement thereof;
  - b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a  
15 nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or a complement thereof under stringent conditions; and
  - c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2.
- 20 9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2.
10. The polypeptide of claim 8 further comprising heterologous amino  
25 acid sequences.
11. An antibody which selectively binds to a polypeptide of claim 8.
12. A method for producing a polypeptide selected from the group  
30 consisting of:
  - a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

10

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

20

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and,

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

30

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

5 19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

- a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and
- b) determining whether the polypeptide binds to the test

10 compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- a) detection of binding by direct detecting of test

15 compound/polypeptide binding;

- b) detection of binding using a competition binding assay;
- c) detection of binding using an assay for 21163-mediated prolyl oligopeptidase activity.

20 21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

25 22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

- a) contacting a polypeptide of claim 8 with a test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the

30 polypeptide.

Input file Fbh21163f1.seq; Output File 21163.trans  
Sequence length 4959

```

GCGTCCGGGACCAGAGGATGAGAGAGCTGTGGCAGGGCTACTGGAACGCTGTCCCTCATGTGTTTAGCCTTCTGTGCTT
CATCTCGAGCTCAGAAGACTAGCGAAGAACGAAAAGTCTTATATATTTTCGAGACTAAAGAACCCAGGAACCCAAGAAGA
CTCCTTGACCACAACAAAGATGGCGAAATCTGCCTGTGCTCGTGCTAGATTGAGGTTACGTAGATTGGCGTGCTGCAA
ACCAGCTCTAGGCGGCTCTGGGTAAGTTGTCTGTTCTGTGGGCTGCGGAACGCAGACTTCGGCTGGACTTGCTGCGGTG
ACACCTGCTCCCTCTGAGAGCTTCAGGTTCTCCGGCTGCCTTCACTGGTTTGTGTCCAGAGCCGGACTGATTCTCTC
M D A F E K
AATTTGCGATCTTCAGCCTGTAAACAAGAAAACGAAAACCCCTTCCAGAAAAC ATG GAT GCA TTT GAA AAA 6
18
V R T K L E T Q P Q E E Y E I I N V E V 26
GTG AGA ACA AAA TTA GAA ACA CAG CCA CAA GAA GAA TAT GAA ATC ATC AAT GTG GAA GTT 78
K H G G F V Y Y Q E G C C L V R S K D E 46
AAA CAT GGT GGT TTT GTT TAT TAC CAA GAA GGT TGT TGC TTG GTT CGT TCC AAA GAT GAA 138
E A D N D N Y E V L F N L E E L K L D Q 66
GAA GCA GAC AAT GAT AAT TAT GAA GTT TTA TTC AAT TTG GAG GAA CTT AAG TTA GAC CAG 198
P F I D C I R V A P D E K Y V A A K I R 86
CCC TTC ATT GAT TGT ATC AGA GTT GCT CCA GAT GAA AAA TAT GTG GCT GCC AAG ATA AGA 258
T E D S E A S T C V I I K L S D Q P V M 106
ACT GAA GAT TCT GAA GCA TCT ACC TGT GTA ATT ATA AAG CTC AGC GAT CAG CCC GTA ATG 318
E A S F P N V S S F E W V K D E E D E D 126
GAA GCT TCT TTC CCG AAT GTG TCC AGT TTT GAA TGG GTA AAG GAC GAG GAA GAT GAA GAT 378
V L F Y T F Q R N L R C H D V Y R A T F 146
GTT TTA TTC TAC ACC TTC CAG AGG AAC CTT CGC TGT CAT GAC GTA TAT CGA GCC ACT TTT 438
G D N K R N E R F Y T E K D P S Y F V F 166
GGT GAT AAC AAA CGT AAT GAA CGC TTT TAC ACA GAA AAA GAC CCA AGC TAC TTT GTT TTC 498
L Y L T K D S R F L T I N I M N K T T S 186
CTT TAT CTT ACA AAA GAC AGT CGT TTC CTC ACC ATA AAT ATT ATG AAC AAG ACT ACT TCT 558
E V W L I D G L S P W D P P V L I Q K R 206
GAA GTG TGG TTG ATA GAT GGC CTG AGC CCT TGG GAC CCA CCA GTA CTT ATC CAG AAG CGA 618
I H G V L Y Y V E H R D D E L Y I L T N 226
ATA CAT GGG GTC CTT TAC TAT GTT GAA CAC AGA GAT GAT GAA TTA TAC ATT CTC ACT AAT 678
V G E P T E F K L M R T A A D T P A I M 246
GTT GGA GAA CCT ACA GAA TTT AAG CTA ATG AGA ACA GCG GCT GAT ACC CCT GCA ATT ATG 738
N W D L F F T M K R N T K V I D L D M F 266
AAT TGG GAT TTA TTT TTT ACA ATG AAG AGA AAT ACA AAA GTG ATA GAC TTG GAC ATG TTT 798
K D H C V L F L K H S N L L Y V N V I G 286
AAG GAT CAC TGT GTT CTA TTT CTG AAG CAC AGC AAT CTC CTT TAT GTT AAT GTG ATT GGT 858
L A D D S V R S L K L P P W A C G F I M 306
CTG GCT GAT GAT TCA GTT CGG TCT CTA AAG CTC CCT CCT TGG GCC TGT GGA TTC ATA ATG 918
D T N S D P K N C P F Q L C S P I R P P 326
GAT ACA AAT TCT GAC CCA AAG AAC TGC CCC TTT CAA CTT TGC TCT CCA ATA CGT CCC CCA 978
K Y Y T Y K F A E G K L F E E T G H E D 346
AAA TAT TAC ACA TAC AAG TTT GCA GAA GGC AAA CTG TTT GAG GAA ACT GGG CAT GAA GAC 1038
P I T K T S R V L R L E A K S K D G K L 366
CCA ATC ACA AAG ACT AGT CGC GTT TTA CGT CTA GAA GCC AAA AGC AAG GAT GGA AAA TTA 1098
V P M T V F H K T D S E D L Q K K P L L 386
GTG CCA ATG ACT GTT TTC CAC AAA ACT GAC TCT GAG GAC TTG CAG AAG AAA CCT CTC TTG 1158

```

FIG. 1A.

V	H	V	Y	G	A	Y	G	M	D	L	K	M	N	F	R	P	E	R	R		406
GTA	CAT	GTA	TAT	GGA	GCT	TAT	GGA	ATG	GAT	TTG	AAA	ATG	AAT	TTC	AGG	CCT	GAG	AGG	CGG		1218
V	L	V	D	D	G	W	I	L	A	Y	C	H	V	R	G	G	G	E	L		426
GTC	CTG	GTG	GAT	GAT	GGA	TGG	ATA	TTA	GCA	TAC	TGC	CAT	GTT	CGA	GGT	GGT	GGT	GAG	TTA		1278
G	L	Q	W	H	A	D	G	R	L	T	K	K	L	N	G	L	A	D	L		446
GGC	CTC	CAG	TGG	CAC	GCT	GAT	GGC	CGC	CTA	ACT	AAA	AAA	CTC	AAT	GGC	CTT	GCT	GAT	TTA		1338
E	A	C	I	K	T	L	H	G	Q	G	F	S	Q	P	S	L	T	T	L		466
GAG	GCT	TGC	ATT	AAG	ACG	CTT	CAT	GGC	CAA	GGC	TTT	TCT	CAG	CCA	AGT	CTA	ACA	ACC	CTG		1398
T	A	F	S	A	G	G	V	L	A	G	A	L	C	N	S	N	P	E	L		486
ACT	GCT	TTC	AGT	GCT	GGA	GGG	GTG	CTT	GCA	GGA	GCA	TTG	TGT	AAT	TCT	AAT	CCA	GAG	CTG		1458
V	R	A	V	T	L	E	A	P	F	L	D	V	L	N	T	M	M	D	T		506
GTG	AGA	GCG	GTG	ACT	TTG	GAG	GCA	CCT	TTC	TTG	GAT	GTT	CTC	AAC	ACC	ATG	ATG	GAC	ACT		1518
T	L	P	L	T	L	E	E	L	E	E	W	G	N	P	S	S	D	E	K		526
ACA	CTT	CCT	CTG	ACA	TTA	GAA	GAA	TTA	GAA	GAA	TGG	GGG	AAT	CCT	TCA	TCT	GAT	GAA	AAA		1578
H	K	N	Y	I	K	R	Y	C	P	Y	Q	N	I	K	P	Q	H	Y	P		546
CAC	AAG	AAC	TAC	ATA	AAA	CGT	TAC	TGT	CCC	TAT	CAA	AAT	ATT	AAA	CCT	CAG	CAT	TAT	CCT		1638
S	I	H	I	T	A	Y	E	N	D	E	R	V	P	L	K	G	I	V	S		566
TCA	ATT	CAC	ATA	ACG	GCA	TAT	GAA	AAC	GAT	GAA	CGG	GTA	CCT	CTG	AAA	GGA	ATT	GTA	AGT		1698
Y	T	E	K	L	K	E	A	I	A	E	H	A	K	D	T	G	E	G	Y		586
TAT	ACT	GAG	AAA	CTC	AAG	GAA	GCC	ATC	GCG	GAG	CAT	GCT	AAG	GAC	ACA	GGT	GAA	GGC	TAT		1758
Q	T	P	N	I	I	L	D	I	Q	P	G	G	N	H	V	I	E	D	S		606
CAG	ACC	CCT	AAT	ATT	ATT	CTA	GAT	ATT	CAG	CCT	GGA	GGC	AAT	CAT	GTA	ATT	GAG	GAT	TCT		1818
H	K	K	I	T	A	Q	I	K	F	L	Y	E	E	L	G	L	D	S	T		626
CAC	AAA	AAG	ATT	ACA	GCC	CAA	ATT	AAA	TTC	CTG	TAC	GAG	GAA	CTT	GGA	CTT	GAC	AGC	ACC		1878
S	V	F	E	D	L	K	K	Y	L	K	F	*									639
AGT	GTT	TTC	GAG	GAT	CTT	AAG	AAA	TAC	CTG	AAA	TTC	TGA									1917

AACACTGCATTCAACTGGGAATTGGAACACACTGAAATATTTTCATAGTCTTACTTCCAATTGAGTTAGCAAAAAAAAA  
ATTAATAACTTGAGACTTTTAAGTTATTAATTTTTTAAAAATGTGCTTCTCCATCTAAATTTTGCTTAGTCTACATCTCA  
CTTGCTTATACTATCTCTCCATTGATGCACATGCCCATTAACCTAGGAAAGTAGTTTTCAAATCATGCTCCTTAGAAG  
GATGTGGAGTAGAGGGAAGGAAGGATTTGGTGATAGCAGAGCTCCAGGCCTCCCTTCCAGTCAGAACAGTTGAGCAGTT  
TACAAATTAGTGTCTGCCTCTTTGCTAGCAAATGCTTTTAGACACTGTGGCAGTGAGTCATCCTCTAATTTCTATGAC  
TGCATTTTAAGGGAAGATAAAATTTCTCCCTTAAATTCGTTAAAGTTTTTGAATAATCTGGGGTCTTAATGTGTT  
CTGGTCATCCCTGATGATGCTATCTGAATAAAGTTATAAGCTCCTATAAGCCATAATTTACTTTTAAACATTTTATTT  
TTTTCAAAACATTTGAGAACCCTTTCTTAAAGCGTTACATTCAAGCTACAGAAATATCGAAGAATTAATGATTGTTTAC  
CAAGCAGCATGCTGTACATGAAGCTATTACAAATGCTTACAATCCCACTGAAATGCCAGTGTCTTCATCTCTTCATAAA  
GGTGCCTAACACGAGGTATACAGTATGTTTCACTGGAATAGCATGCTCGATTGGAACAAAGCATCTATCTCTGA  
AAGCTGTTTGGCGATGAAGGAGATTCTTCGTGTTGTGTTCAAAGATGAGTCCCTCTCCCTTGTCCAGAAAAATGCCACT  
TGTATCAACTTTACTGCCTTTGTGCGCAGAATTGGTACTTAACCTTATCTTATTTTAGCGGAAGGCCGAAATCATA  
TTATGTAGATTTAACAGTGTGATTCTCCAAAATTCAGAACACGATAAAGATTCTGTGCGATGCCATCCAGCTCTCTTG  
TGTACACAACATAGTGGCTGTCAATTCTCAATGGCAAACAGCCCTGTTGAGGAGTAGCTCATGGCATGAAGTAG  
ACTTAAATCTTGATATAACTTCAAAGCCGATCTGGGCTGAGTCTTTTGGACCTATTTTTTAAAAAGTATTTACGTAA  
GTGTTTGATTCTAAGAATTGTTTGTAAGTATTTTAAATATATTGTAAGGAGTTATTTACCCAAAACACTTGCTCCAATT

FIG. 1B.

TTGCCCCTTATAATTGCCAAATTGTAAGCATCAATAAGTAGGTAAGAACAATTTATATAAAAACGATAGAAATGACAA  
ATTCGGGGTTTCGGCTTGTCCGGGAGTCAATAAGTACGCACAGTGCTCTGCTACATTGTAGAGTTTCTGTAGAGATCAA  
ATTTGACTCCACTTTAGGAGTCCCAAAGCAAATGTCCATGTCTAAGATGAATATTTAACTTGCATAGTCATTCTGTGCT  
ATATTGTAAC TGCCAGATGGCCAGAAAGAAAGGCAACAGTGGACTCAGACTTCTGAGGAATTTGGGTTTGTTCCTTTG  
TAGACTAATGTGTAGGTTGCTGTTGTGCGAAGATCGTGTAAC TTTAGCAGACATGTATTTCTTGCACAGCTAATAGAAG  
ACAAAGTTGAAAAAAGGATGCAAAATAAAAAGCTGCCTAAGGTGAAAGTTAGAAATGTAGACTTTTTTTTACCATAA  
TAGTATGTGTTCATTGAAGATGATTTGGGTTTATTTTACAGCTATATAAAACATAATTTGATGATGTACTTCTAACCTT  
TCAAGCATT TTTCTGTTATTGACTATATAATATAGCCTCCATAAATGTTTTTAATGACAATATCTGTGAACGGTTGTA  
CCATACTCAGCCATGCCCTTTTCATTTTGACGATAGTGTTCCTAATATTTTGTATTTTATTCCTCCCTCCCCCA TTTTGT  
TATTACTTAAGATAGATTATCAGAAAGACAGTTACTTTGTCAAAGAGTATGGGCAC TTGATACATAATGCCAAATTATT  
CTTCATAAGAGCTGTTGCCAAATCAGTGATAATGTTCA TTTAATTGTATTCTTGCCAGCCATGTTTACTGGGGTGATAG  
TTGTATTGTGGTTGTTATTGTTCTTTAGGGGTAGGTTCCCAATATGTGGTCTTTAAATAATTATCTAATGGTGT TTA  
AAAGATGTTTATTCGTTTGTTCAGGTACAAAGATATTTATGATACATGTATGACTTGTCTAAGTTATTAACATTTTCTC  
TAGCCTTAGGTAATGCATGAAAGCACATGTTTCAGTGCCACTCACATAAGAAGTGCCCGTAAGTGT TAGCTATTATTG  
TCTACTTGAGTTACTACTTTCTAAAAGTATGTTGAAGTCTTTTCTGTAATTGCAGATTTGTTGATTTTGCATTTGAGT  
ATTTTCTATATTTTGAAGCTGTTAGATGCATAGTCATGATTTTGGTGGAATGTTTATCAATTTTGA AAAATTGCCTT  
TGTCTCATATAATGCTTTTCATATTGAAC TATATTTTGTCTGCTATTAAATACTTCCAAGCCTG

FIG. 1C.

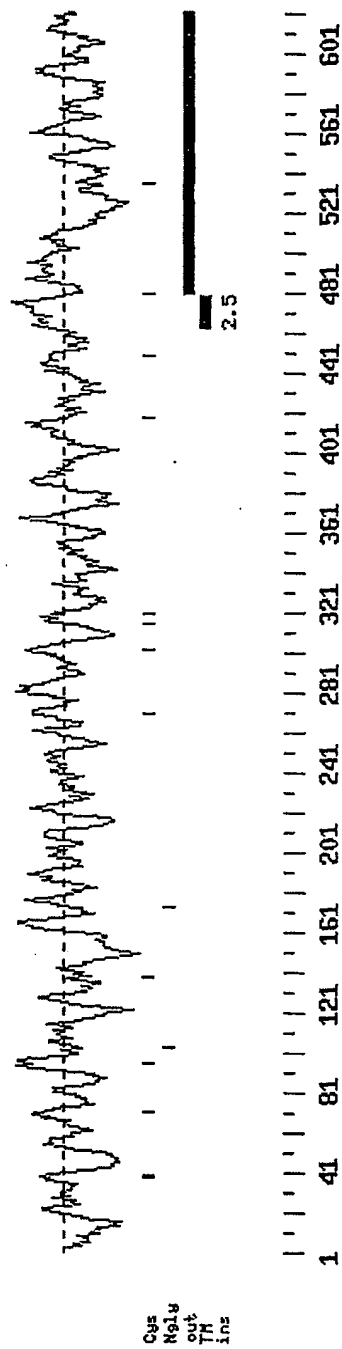


FIG. 2.

## Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

-----

HMM file:                   /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file:             /prod/ddm/wspace/orfanal/oa-script.12506.seq

-----

Query: Fbh21163f1

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Peptidase_S9	Prolyl oligopeptidase family	45.1	1.1e-12	1
trans_reg_C	Transcriptional regulatory protein, C	4.4	6.7	1
Peptidase_S9_N	Prolyl oligopeptidase, N-terminal beta	-52.5	4.8e-07	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Peptidase_S9_N	1/1	1	342 [.	1	464 []	-52.5	4.8e-07
trans_reg_C	1/1	438	457 ..	56	75 .]	4.4	6.7
Peptidase_S9	1/1	397	475 ..	1	81 []	45.1	1.1e-12

FIG. 3A.



## Alignments of top-scoring domains:

Peptidase\_S9\_N: domain 1 of 1, from 1 to 342: score -52.5, E = 4.8e-07

```

      *->yPptprdesvvdelhGdvasvraDPYrWLeDddRtnpevlawleaEN
      ++ + v ++
Fbh21163f1 1 -----MDAFEKV-----RT----- 9
      kytqaflaqlkplrekifeeltalineddisaPfrkGgyyYYfrntggkn
      l+ p++e ++ i + ++Gg+ YY+
Fbh21163f1 10 -----KLETQ-PQEE-----YBIINVEVKHGGFVYQEG----- 37
      ypvlcRrdalssyrtegetisAlnadfeeEeVLDPntlaedgdFtvalg
      c + ++ ++e+ d + eVl + +l d F + +
Fbh21163f1 38 ---CCLVRSK-----DEEA-----DNDNYEVLNLEELKLDQPPIDCIR 73
      vfavSpDgnllAYsvdpgGsDwytIrfkDletgrellpDelegvkfsgsi
      v+ pD +++A ++ s++ t +++l+++ ++ + + ++ s+ +
Fbh21163f1 74 VA---PDEKYVAAKIRTEDSEASTCVIIKLSAQ--PVMEA-SFPNVSS-F 116
      vWapD...gkgfFYtrydepqrgsDddGtetatnlpkhVwrHrLGtpqsg
      +W +D++++++FYt + r h V+r G + +
Fbh21163f1 117 EWVKDeedEDVLFYTFQRN-LR-----CHDVYRATFGDNKRN 152
      DelvfeepddpfwyvgaevSrDgkYlfissgsgtdpenelyvldlgseai
      e ++e+d+++ +v+++++D ++l i++++t+ +e++++d s
Fbh21163f1 153 -ERFYTEKDPSY-FVFLYLTKDSRFLTINIMNKT--SEVWLIDGLS--- 195
      clpagnphaplrllvprfdgvyvyyvehhkgGgdrfyiltNsdgApNfrLv
      p +p+ +++r++gv yyveh+ +d++yiltN f+L+
Fbh21163f1 196 -----PWDPPVLIQKRIHGVLYVEHR---DDELYILTNGEPTFKLM 236
      rvdlnpdpadwktvivehredvllleeiavfgnyLvvsyrkdglslrlqvi
      r+ ++p+ +w +++ + ++ + + ++f+++ v++ + l + vi
Fbh21163f1 237 RTAADTPAIMNWD-LFFTMKRNTKVIDLDMFKDHCVLFLKHSNLLYVNV 285
      dlgakdggfgggaelgkvsfeepplpgvgsvssasgeydsdelrysfsSflt
      l + + +s++ p+ ++g + ++ d ++ +++ +S +
Fbh21163f1 286 GLA-----DDSVRSCLKLPWACGFIMDTN--SDPKNCPFQLCSPIR 324
      Pstiydlldtgerellkdr<-*
      P+ y + ++ g+ l+ ++
Fbh21163f1 325 PPKYITYKFAEGK--LFEET 342

```

trans\_reg\_C: domain 1 of 1, from 438 to 457: score 4.4, E = 6.7

```

      *->kKLeddpaspryIkTVRGvG<-*
      kKL+ + + IkT++G+G
Fbh21163f1 438 KKLNGLEADLEACIKTLHGQG 457

```

Peptidase\_S9: domain 1 of 1, from 397 to 475: score 45.1, E = 1.1e-12

```

      *->sltpnFsvsvasllnhrGgiyAvvdiRGgGeyGqkwheagtrrlkkn
      nF + l++ G i+A ++RGgGe+G +wh +g+ +k n
Fbh21163f1 397 -LKMNFRPERRVLVD-DGWILAYCHVRGGGELGLQWHADGRLTKKLN 441
      efnDfiaAAeylsklGYtspkriaifGgSnGGIL<-*
      D a l G+ p++ ++ S+GG L
Fbh21163f1 442 GLADLEACIKTLHGQGFSPSLTTLTAFSAGGVL 475

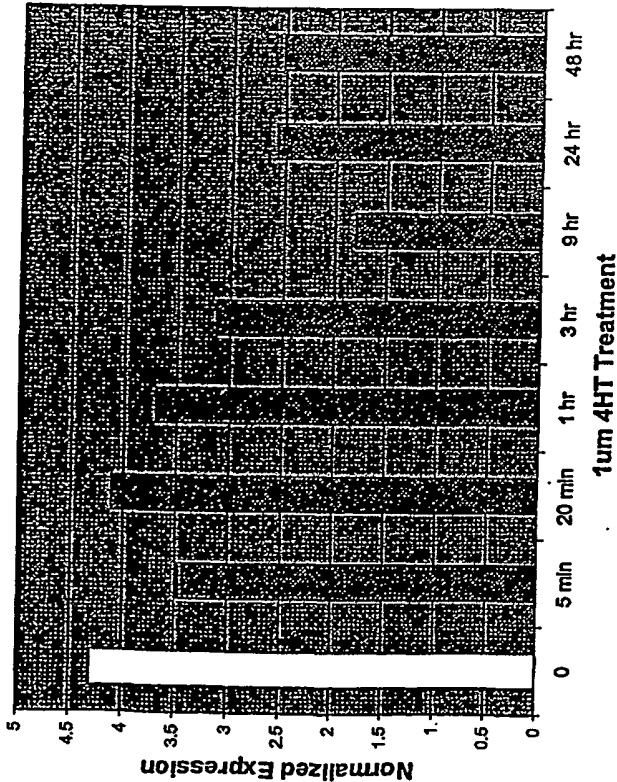
```

//

FIG. 3B.

Tumor Suppressor Induced Down-regulation of 21163 Expression

a) p53 induced down-regulation (TxP, MPGv5)



b) p16 induced down-regulation (Taqman)

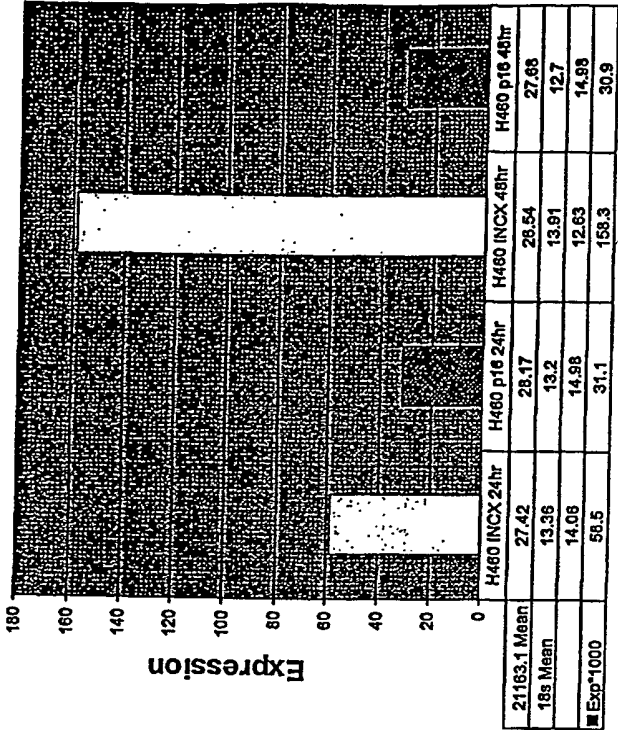


FIG. 4.

Expression of 21163 in Normal and Diseased Tissues

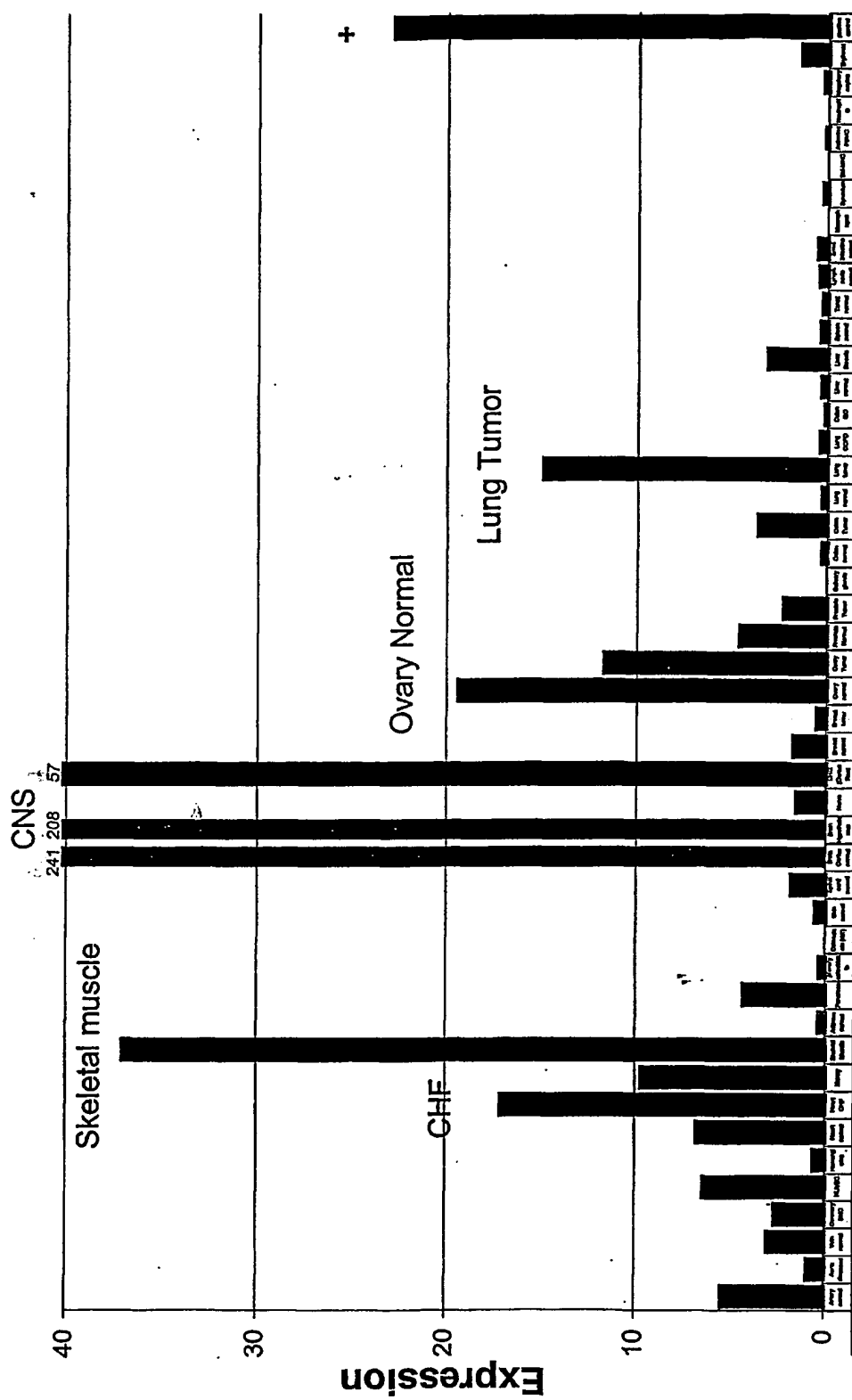


FIG. 5.

## SEQUENCE LISTING

<110> Hunter, John Joseph  
Kapeller-Libermann, Rosana

<120> 21163, A Novel Human Prolyl  
Oligopeptidase and Uses Therefor

<130> 035800/22198

<150> 60/257,736

<151> 2000-12-22

<160> 6

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 4959

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (451)...(2364)

<400> 1

```

gcgtccggga ccagaggatg agagagctgt ggcagggcta ctggaacgct gtccctcatt 60
gttttagcct tctgtgcttc atctcgagct cagaagacta gcgaagaacg aaaagtctta 120
tatatttcga gactaaagaa cccaggaacc caagaagact ccttgaccac aacaaagatg 180
gcgaaatctg cctgttgctc gtgctagatt gaggttacgt agattggcgt gctgcaaacc 240
agctctaggc ggctctgggt aagttgtcgt tctgtgggct gcggaacgca gacttcggct 300
ggacttgccct gcggtgacac ctgctcccct ctgagagctt caggttctcc ggctgcctt 360
cactggtttg tgtccagagc cggactgatt ctctcaattt gcgatcttca gcctgttaaa 420
caagaaaacg aaaaaccctt tccagaaaac atg gat gca ttt gaa aaa gtg aga 474
                                Met Asp Ala Phe Glu Lys Val Arg
                                1                               5

aca aaa tta gaa aca cag cca caa gaa gaa tat gaa atc atc aat gtg 522
Thr Lys Leu Glu Thr Gln Pro Gln Glu Glu Tyr Glu Ile Ile Asn Val
    10                               15                               20

gaa gtt aaa cat ggt ggt ttt gtt tat tac caa gaa ggt tgt tgc ttg 570
Glu Val Lys His Gly Gly Phe Val Tyr Tyr Gln Glu Gly Cys Cys Leu
    25                               30                               35                               40

gtt cgt tcc aaa gat gaa gaa gca gac aat gat aat tat gaa gtt tta 618
Val Arg Ser Lys Asp Glu Glu Ala Asp Asn Asp Asn Tyr Glu Val Leu
                                45                               50                               55

ttc aat ttg gag gaa ctt aag tta gac cag ccc ttc att gat tgt atc 666
Phe Asn Leu Glu Glu Leu Lys Leu Asp Gln Pro Phe Ile Asp Cys Ile
                                60                               65                               70

aga gtt gct cca gat gaa aaa tat gtg gct gcc aag ata aga act gaa 714
Arg Val Ala Pro Asp Glu Lys Tyr Val Ala Ala Lys Ile Arg Thr Glu
                                75                               80                               85

gat tct gaa gca tct acc tgt gta att ata aag ctc agc gat cag ccc 762
Asp Ser Glu Ala Ser Thr Cys Val Ile Ile Lys Leu Ser Asp Gln Pro
                                90                               95                               100

gta atg gaa gct tct ttc ccg aat gtg tcc agt ttt gaa tgg gta aag 810
Val Met Glu Ala Ser Phe Pro Asn Val Ser Ser Phe Glu Trp Val Lys
    105                               110                               115                               120

```

gac gag gaa gat gaa gat gtt tta ttc tac acc ttc cag agg aac ctt	858
Asp Glu Glu Asp Glu Asp Val Leu Phe Tyr Thr Phe Gln Arg Asn Leu	
125 130 135	
cgc tgt cat gac gta tat cga gcc act ttt ggt gat aac aaa cgt aat	906
Arg Cys His Asp Val Tyr Arg Ala Thr Phe Gly Asp Asn Lys Arg Asn	
140 145 150	
gaa cgc ttt tac aca gaa aaa gac cca agc tac ttt gtt ttc ctt tat	954
Glu Arg Phe Tyr Thr Glu Lys Asp Pro Ser Tyr Phe Val Phe Leu Tyr	
155 160 165	
ctt aca aaa gac agt cgt ttc ctc acc ata aat att atg aac aag act	1002
Leu Thr Lys Asp Ser Arg Phe Leu Thr Ile Asn Ile Met Asn Lys Thr	
170 175 180	
act tct gaa gtg tgg ttg ata gat ggc ctg agc cct tgg gac cca cca	1050
Thr Ser Glu Val Trp Leu Ile Asp Gly Leu Ser Pro Trp Asp Pro Pro	
185 190 195 200	
gta ctt atc cag aag cga ata cat ggg gtc ctt tac tat gtt gaa cac	1098
Val Leu Ile Gln Lys Arg Ile His Gly Val Leu Tyr Tyr Val Glu His	
205 210 215	
aga gat gat gaa tta tac att ctc act aat gtt gga gaa cct aca gaa	1146
Arg Asp Asp Glu Leu Tyr Ile Leu Thr Asn Val Gly Glu Pro Thr Glu	
220 225 230	
ttt aag cta atg aga aca gcg gct gat acc cct gca att atg aat tgg	1194
Phe Lys Leu Met Arg Thr Ala Ala Asp Thr Pro Ala Ile Met Asn Trp	
235 240 245	
gat tta ttt ttt aca atg aag aga aat aca aaa gtg ata gac ttg gac	1242
Asp Leu Phe Phe Thr Met Lys Arg Asn Thr Lys Val Ile Asp Leu Asp	
250 255 260	
atg ttt aag gat cac tgt gtt cta ttt ctg aag cac agc aat ctc ctt	1290
Met Phe Lys Asp His Cys Val Leu Phe Leu Lys His Ser Asn Leu Leu	
265 270 275 280	
tat gtt aat gtg att ggt ctg gct gat gat tca gtt cgg tct cta aag	1338
Tyr Val Asn Val Ile Gly Leu Ala Asp Asp Ser Val Arg Ser Leu Lys	
285 290 295	
ctc cct cct tgg gcc tgt gga ttc ata atg gat aca aat tct gac cca	1386
Leu Pro Pro Trp Ala Cys Gly Phe Ile Met Asp Thr Asn Ser Asp Pro	
300 305 310	
aag aac tgc ccc ttt caa ctt tgc tct cca ata cgt ccc cca aaa tat	1434
Lys Asn Cys Pro Phe Gln Leu Cys Ser Pro Ile Arg Pro Pro Lys Tyr	
315 320 325	
tac aca tac aag ttt gca gaa ggc aaa ctg ttt gag gaa act ggg cat	1482
Tyr Thr Tyr Lys Phe Ala Glu Gly Lys Leu Phe Glu Glu Thr Gly His	
330 335 340	
gaa gac cca atc aca aag act agt cgc gtt tta cgt cta gaa gcc aaa	1530
Glu Asp Pro Ile Thr Lys Thr Ser Arg Val Leu Arg Leu Glu Ala Lys	
345 350 355 360	
agc aag gat gga aaa tta gtg cca atg act gtt ttc cac aaa act gac	1578
Ser Lys Asp Gly Lys Leu Val Pro Met Thr Val Phe His Lys Thr Asp	
365 370 375	
tct gag gac ttg cag aag aaa cct ctc ttg gta cat gta tat gga gct	1626

Ser	Glu	Asp	Leu	Gln	Lys	Lys	Pro	Leu	Leu	Val	His	Val	Tyr	Gly	Ala		
			380					385					390				
tat	gga	atg	gat	ttg	aaa	atg	aat	ttc	agg	cct	gag	agg	cgg	gtc	ctg	1674	
Tyr	Gly	Met	Asp	Leu	Lys	Met	Asn	Phe	Arg	Pro	Glu	Arg	Arg	Val	Leu		
		395					400					405					
gtg	gat	gat	gga	tggt	ata	tta	gca	tac	tgc	cat	gtt	cga	ggg	ggg	ggg	1722	
Val	Asp	Asp	Gly	Trp	Ile	Leu	Ala	Tyr	Cys	His	Val	Arg	Gly	Gly	Gly		
		410				415					420						
gag	tta	ggc	ctc	cag	tggt	cac	gct	gat	ggc	cgc	cta	act	aaa	aaa	ctc	1770	
Glu	Leu	Gly	Leu	Gln	Trp	His	Ala	Asp	Gly	Arg	Leu	Thr	Lys	Lys	Leu		
		425			430					435					440		
aat	ggc	ctt	gct	gat	tta	gag	gct	tgc	att	aag	acg	ctt	cat	ggc	caa	1818	
Asn	Gly	Leu	Ala	Asp	Leu	Glu	Ala	Cys	Ile	Lys	Thr	Leu	His	Gly	Gln		
			445					450						455			
ggc	ttt	tct	cag	cca	agt	cta	aca	acc	ctg	act	gct	ttc	agt	gct	gga	1866	
Gly	Phe	Ser	Gln	Pro	Ser	Leu	Thr	Thr	Leu	Thr	Ala	Phe	Ser	Ala	Gly		
			460					465					470				
ggg	gtg	ctt	gca	gga	gca	ttg	tgt	aat	tct	aat	cca	gag	ctg	gtg	aga	1914	
Gly	Val	Leu	Ala	Gly	Ala	Leu	Cys	Asn	Ser	Asn	Pro	Glu	Leu	Val	Arg		
		475				480						485					
gcg	gtg	act	ttg	gag	gca	cct	ttc	ttg	gat	gtt	ctc	aac	acc	atg	atg	1962	
Ala	Val	Thr	Leu	Glu	Ala	Pro	Phe	Leu	Asp	Val	Leu	Asn	Thr	Met	Met		
		490				495					500						
gac	act	aca	ctt	cct	ctg	aca	tta	gaa	gaa	tta	gaa	gaa	tggt	ggg	aat	2010	
Asp	Thr	Thr	Leu	Pro	Leu	Thr	Leu	Glu	Glu	Leu	Glu	Glu	Trp	Gly	Asn		
				510						515				520			
cct	tca	tct	gat	gaa	aaa	cac	aag	aac	tac	ata	aaa	cgt	tac	tgt	ccc	2058	
Pro	Ser	Ser	Asp	Glu	Lys	His	Lys	Asn	Tyr	Ile	Lys	Arg	Tyr	Cys	Pro		
				525				530						535			
tat	caa	aat	att	aaa	cct	cag	cat	tat	cct	tca	att	cac	ata	acg	gca	2106	
Tyr	Gln	Asn	Ile	Lys	Pro	Gln	His	Tyr	Pro	Ser	Ile	His	Ile	Thr	Ala		
			540					545					550				
tat	gaa	aac	gat	gaa	cgg	gta	cct	ctg	aaa	gga	att	gta	agt	tat	act	2154	
Tyr	Glu	Asn	Asp	Glu	Arg	Val	Pro	Leu	Lys	Gly	Ile	Val	Ser	Tyr	Thr		
		555				560						565					
gag	aaa	ctc	aag	gaa	gcc	atc	gcg	gag	cat	gct	aag	gac	aca	ggg	gaa	2202	
Glu	Lys	Leu	Lys	Glu	Ala	Ile	Ala	Glu	His	Ala	Lys	Asp	Thr	Gly	Glu		
		570				575					580						
ggc	tat	cag	acc	cct	aat	att	att	cta	gat	att	cag	cct	gga	ggc	aat	2250	
Gly	Tyr	Gln	Thr	Pro	Asn	Ile	Ile	Leu	Asp	Ile	Gln	Pro	Gly	Gly	Asn		
				590						595				600			
cat	gta	att	gag	gat	tct	cac	aaa	aag	att	aca	gcc	caa	att	aaa	ttc	2298	
His	Val	Ile	Glu	Asp	Ser	His	Lys	Lys	Ile	Thr	Ala	Gln	Ile	Lys	Phe		
				605					610					615			
ctg	tac	gag	gaa	ctt	gga	ctt	gac	agc	acc	agt	gtt	ttc	gag	gat	ctt	2346	
Leu	Tyr	Glu	Glu	Leu	Gly	Leu	Asp	Ser	Thr	Ser	Val	Phe	Glu	Asp	Leu		
			620					625					630				
aag	aaa	tac	ctg	aaa	ttc	tgaaacactg	cattcaactg	ggaattggaa								2394	
Lys	Lys	Tyr	Leu	Lys	Phe												
			635														

```

acacactgaa atatttcata gtcttacttc caattgagtt agcaaaaaaa aaattaataa 2454
cttgagactt ttaagttatt aattttttta aatgtgcttc tccatctaaa ttttgcttag 2514
tctacatctc acttgcttat actattctct ccattgatgc acatgcccat taacctagga 2574
aagtagtttt caaatcatgc tccttagaag gatgtggagt agaggggaagg gaaggattgg 2634
tgatagcaga gctccaggcc tcccttccag tcagaacagt tgagcagttt acaaattagt 2694
gtcctgcctc tttgctagca aatgctttta gacactgtgg cagtgaagca tcctctaatt 2754
tctatgactg cattttaagg gaaaagataa aattcttccc cttaaaattc gttaaagttt 2814
ttgaataatc tggggtccta atgtgttctg gtcatccctg attgatgcta tctgaataaa 2874
gttataagct cctataagcc ataatttact tttaaacatt ttattttttt caaaacattt 2934
gagaaccttt cttaaagcgg ttacattcaa gctacagaaa tatcgaagaa ttaatgattg 2994
ttcaccaagc agcatgctgt acatgaagct attacaaatg cttacaatcc cactgaaatg 3054
ccagtgtctt catctcttca taaaggtgcc taacacgagg tatacagtat gttcagtaca 3114
ctggaatagc atgctcgatt ggaacaaaag catctatctc tgaagagctg ttggcgatga 3174
aggagattct tcgtgttgtg ttcaaagatg agtccctctc ccttgtccag aaaaatgcc 3234
cttgatatca ctttactgcc tttgtcggca gaattggtac ttaaccttat tcttatttta 3294
gcgggaaggc ccgaaatcat attatgtaga tttaacagtg ttgattctcc aaaattcaga 3354
accacgataa agattctgtc gatgccatcc agctctcttg tgtacacaac atagtggctg 3414
tcattctcca aatggcaaaa ccagccccctg ttgaggagta gctcattggc atgaagtaga 3474
cttaaattct gatataactt caaagccgat ctgggctgag tcttttggac ctattttttt 3534
aaaaaagtat ttacgtaagt gtttgattct aagaattgtt tgtaagtatt tttaatatat 3594
tgtaaggagt tatttaccce aaacacttgc tccaattttg ccccttataa ttgccaaatt 3654
gtaagcatca ataagtaggt aagaacaatt tatataaaaa ctgatagaaa tgacaaattc 3714
ggggtttcgg cttgtccggg agtcaataag tacgcacagt gctctgtctac attgtagagt 3774
ttctgtagag atcaaatttg actccacttt aggagtccca aagcaaatgt ccatgtctaa 3834
gatgaatatt taacttgcac agtcattctg tgctatattg taactgccag atggccagaa 3894
agaaggcaac agtggactca gacttctgag gaatttgggt ttgttcccct ttgtagacta 3954
atgtgtaggt tgctgttgtg cgaagatcgt gtaactttag cagacatgta tttcttgcac 4014
agctaataga agacaaagt gaaaaaaagg atgcaaaata aaaagctgcc taaggtgaaa 4074
gttagaaaatt gtagactttt ttttaccata atagtatgtg ttcattgaag atgatttggg 4134
tttatttttac agctatataa aacataattt gatgatgtac ttctaacctt tcaagcattt 4194
tctgttatgg actatataat atagcctcca taaatgtttt taatgacaat attctgttga 4254
acggttgtac catactcagc catgcccttt cattttgacg atagtgtttc taatattttg 4314
tattttttatt cccctcccc cacttttggta ttacttaaga tagattatca gaaagacagt 4374
tacttttgta aagagtatgg gcacttgata cataatgcc aattattctt cataagagct 4434
gttgccaaata cagtgataat gttcatttaa ttgtattctt gccagccatg tttactgggg 4494
tgatagttgt tattgtggtt gttattgttc tttaggggta ggttcccaat atgtggtctt 4554
taaataatta tctaattgtg tttaaaaaga tgtttattct gtttgtcagg tacaagata 4614
tttatgatac atgtatgact tgtctaagtt attaacattt tctctagcct taggtaatgc 4674
atgaaagcac atgtttcagt gccactcaca taagaagtgc ccggtaaagt ttagctatta 4734
ttgtctactt gagtactac tttctaaaag tatgttgaag tctttttctg taattgcaga 4794
tttgttgatt ttgcatttga gtattttcta tattttgaag ctgttagatg catagtcatg 4854
atttttggtg gaatgtttta tcaatttttg aaaattgcct ttgtctcata taatgctttt 4914
catattgaac tataatttgt ctgctattaa atacttccaa gcctg 4959

```

&lt;210&gt; 2

&lt;211&gt; 638

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

Met Asp Ala Phe Glu Lys Val Arg Thr Lys Leu Glu Thr Gln Pro Gln
1           5           10          15
Glu Glu Tyr Glu Ile Ile Asn Val Glu Val Lys His Gly Gly Phe Val
20          25          30
Tyr Tyr Gln Glu Gly Cys Cys Leu Val Arg Ser Lys Asp Glu Glu Ala
35          40          45
Asp Asn Asp Asn Tyr Glu Val Leu Phe Asn Leu Glu Glu Leu Lys Leu
50          55          60
Asp Gln Pro Phe Ile Asp Cys Ile Arg Val Ala Pro Asp Glu Lys Tyr
65          70          75          80
Val Ala Ala Lys Ile Arg Thr Glu Asp Ser Glu Ala Ser Thr Cys Val
85          90          95
Ile Ile Lys Leu Ser Asp Gln Pro Val Met Glu Ala Ser Phe Pro Asn
100         105         110
Val Ser Ser Phe Glu Trp Val Lys Asp Glu Glu Asp Glu Asp Val Leu

```

	115		120		125	
Phe	Tyr Thr Phe Gln Arg Asn Leu Arg Cys His Asp Val Tyr Arg Ala					
130		135		140		
Thr	Phe Gly Asp Asn Lys Arg Asn Glu Arg Phe Tyr Thr Glu Lys Asp					
145		150		155		160
Pro	Ser Tyr Phe Val Phe Leu Tyr Leu Thr Lys Asp Ser Arg Phe Leu					
	165		170			175
Thr	Ile Asn Ile Met Asn Lys Thr Thr Ser Glu Val Trp Leu Ile Asp					
	180		185			190
Gly	Leu Ser Pro Trp Asp Pro Pro Val Leu Ile Gln Lys Arg Ile His					
	195		200			205
Gly	Val Leu Tyr Tyr Val Glu His Arg Asp Asp Glu Leu Tyr Ile Leu					
210		215		220		
Thr	Asn Val Gly Glu Pro Thr Glu Phe Lys Leu Met Arg Thr Ala Ala					
225		230		235		240
Asp	Thr Pro Ala Ile Met Asn Trp Asp Leu Phe Phe Thr Met Lys Arg					
	245		250			255
Asn	Thr Lys Val Ile Asp Leu Asp Met Phe Lys Asp His Cys Val Leu					
	260		265			270
Phe	Leu Lys His Ser Asn Leu Leu Tyr Val Asn Val Ile Gly Leu Ala					
275		280		285		
Asp	Asp Ser Val Arg Ser Leu Lys Leu Pro Pro Trp Ala Cys Gly Phe					
290		295		300		
Ile	Met Asp Thr Asn Ser Asp Pro Lys Asn Cys Pro Phe Gln Leu Cys					
305		310		315		320
Ser	Pro Ile Arg Pro Pro Lys Tyr Tyr Thr Tyr Lys Phe Ala Glu Gly					
	325		330			335
Lys	Leu Phe Glu Thr Gly His Glu Asp Pro Ile Thr Lys Thr Ser					
	340		345			350
Arg	Val Leu Arg Leu Glu Ala Lys Ser Lys Asp Gly Lys Leu Val Pro					
	355		360			365
Met	Thr Val Phe His Lys Thr Asp Ser Glu Asp Leu Gln Lys Lys Pro					
370		375		380		
Leu	Leu Val His Val Tyr Gly Ala Tyr Gly Met Asp Leu Lys Met Asn					
385		390		395		400
Phe	Arg Pro Glu Arg Arg Val Leu Val Asp Asp Gly Trp Ile Leu Ala					
	405		410			415
Tyr	Cys His Val Arg Gly Gly Gly Glu Leu Gly Leu Gln Trp His Ala					
	420		425			430
Asp	Gly Arg Leu Thr Lys Lys Leu Asn Gly Leu Ala Asp Leu Glu Ala					
	435		440			445
Cys	Ile Lys Thr Leu His Gly Gln Gly Phe Ser Gln Pro Ser Leu Thr					
450		455		460		
Thr	Leu Thr Ala Phe Ser Ala Gly Gly Val Leu Ala Gly Ala Leu Cys					
465		470		475		480
Asn	Ser Asn Pro Glu Leu Val Arg Ala Val Thr Leu Glu Ala Pro Phe					
	485		490			495
Leu	Asp Val Leu Asn Thr Met Met Asp Thr Thr Leu Pro Leu Thr Leu					
	500		505			510
Glu	Glu Leu Glu Glu Trp Gly Asn Pro Ser Ser Asp Glu Lys His Lys					
	515		520			525
Asn	Tyr Ile Lys Arg Tyr Cys Pro Tyr Gln Asn Ile Lys Pro Gln His					
530		535		540		
Tyr	Pro Ser Ile His Ile Thr Ala Tyr Glu Asn Asp Glu Arg Val Pro					
545		550		555		560
Leu	Lys Gly Ile Val Ser Tyr Thr Glu Lys Leu Lys Glu Ala Ile Ala					
	565		570			575
Glu	His Ala Lys Asp Thr Gly Glu Gly Tyr Gln Thr Pro Asn Ile Ile					
	580		585			590
Leu	Asp Ile Gln Pro Gly Gly Asn His Val Ile Glu Asp Ser His Lys					
	595		600			605
Lys	Ile Thr Ala Gln Ile Lys Phe Leu Tyr Glu Glu Leu Gly Leu Asp					
610		615		620		
Ser	Thr Ser Val Phe Glu Asp Leu Lys Lys Tyr Leu Lys Phe					
625		630		635		



<210> 3  
 <211> 1914  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 atgggatgcat ttgaaaaagt gagaacaaaa ttagaaacac agccacaaga agaatatgaa 60  
 atcatcaatg tggaaagttaa acatgggtggt tttgtttatt accaagaagg ttgttgcttg 120  
 gttcgtttcca aagatgaaga agcagacaat gataattatg aagttttatt caatttggag 180  
 gaacttaagt tagaccagcc cttcattgat tgtatcagag ttgctccaga tgaaaaatat 240  
 gtggctgcca agataagaac tgaagattct gaagcatcta cctgtgtaat tataaagctc 300  
 agcgatcagc ccgtaatgga agcttctttc ccgaatgtgt ccagttttga atgggtaaag 360  
 gacgaggaag atgaagatgt tttattctac accttccaga ggaaccttcg ctgtcatgac 420  
 gtatatcgag ccacttttgg tgataacaaa cgtaatgaac gcttttacac agaaaaagac 480  
 ccaagctact ttgttttcct ttatcttaca aaagacagtc gtttcctcac cataaatatt 540  
 atgaacaaga ctacttctga agtgtggttg atagatggcc tgagcccttg ggaccaccca 600  
 gtacttatcc agaagcgaat acatggggtc ctttactatg ttgaacacag agatgatgaa 660  
 ttatacattc tcactaatgt tggagaacct acagaattta agctaattgag aacagcggct 720  
 gataccctg caattatgaa ttgggattta tttttacaa tgaagagaaa taaaaagt 780  
 atagacttgg acatgtttta ggatcactgt gttctatttc tgaagcacag caatctcctt 840  
 tatgttaatg tgattggtct ggctgatgat tcagttcggg ctctaaagct ccctccttgg 900  
 gcctgtggat tcataatgga tacaaattct gacccaaaga actgcccctt tcaactttgc 960  
 tctccaatag gtccccaaa atattacaca tacaagtttg cagaaggcaa actgtttgag 1020  
 gaaactgggc atgaagaccc aatcacaaag actagtcgcg ttttacgtct agaagccaaa 1080  
 agcaaggatg gaaaattagt gccaatgact gttttccaca aaactgactc tgaggacttg 1140  
 cagaagaac ctctcttggg acatgtatat ggagcttatg gaatggattt gaaaatgaat 1200  
 ttcaggcctg agaggcgggt cctggtggat gatggatgga tattagcata ctgccatgtt 1260  
 cgaggtgggt gtgagttagg cctccagtgg cagctgatg gccgcctaac taaaaaactc 1320  
 aatggccttg ctgatttaga ggcttgcat aagacgcttc atggccaagg cttttctcag 1380  
 ccaagtctaa caacctgac tgctttcagt gctggagggt tgcttgacag agcatttgtt 1440  
 aattctaatc cagagctggt gagagcgggt actttggagg cacctttctt ggatgttctc 1500  
 aacaccatga tggacactac acttctctg acattagaag aattagaaga atgggggaat 1560  
 accttcacatg atgaaaaaca caagaactac ataaaacgtt actgtcccta tcaaatatt 1620  
 aaacctcagc attatctctc aattcacata acggcatatg aaaacgatga acgggtacct 1680  
 ctgaaaggaa ttgtaagtta tactgagaaa ctcaaggaag ccacgcgga gcatgctaag 1740  
 gacacaggtg aaggctatca gaccctaatt attattctag atattcagcc tggaggcaat 1800  
 catgtaatg aggattctca caaaaagatt acagcccaa ttaaattcct gtacgaggaa 1860  
 ctgtgacttg acagcaccag tgttttcgag gatcttaaga aatacctgaa attc 1914

<210> 4  
 <211> 55  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> PFAM consensus sequence for Prolyl Oligopeptidase  
 Family

<400> 4  
 Pro Gly Asp Pro Tyr Trp Leu Asp Arg Glu Asn Pro Gly Tyr Tyr Arg  
 1 5 10 15  
 Ala Glu Val Leu Asp Asn Phe Ser Asp Ala Tyr Gly Asp Ile Asp Asp  
 20 25 30  
 Trp Asp Phe Tyr Asp Gly Val His Leu Gly Asp Ser Asp Tyr Gly Thr  
 35 40 45  
 Asn Ala Asn Leu Leu Ser Pro  
 50 55

<210> 5  
 <211> 8  
 <212> PRT  
 <213> Artificial Sequence

<220>

<223> PFAM consensus sequence for Transcriptional  
Regulatory Protein, C

<400> 5

Lys Leu Ile Thr Val Arg Gly Gly  
1 5

<210> 6

<211> 17

<212> PRT

<213> Artificial Sequence

<220>

<223> PFAM consensus sequence for Prolyl Oligopeptidase,  
N-terminal Beta-Propeller Domain

<400> 6

Phe Gly Ala Arg Gly Gly Gly Asp Ala Ala Gly Tyr Gly Ser Gly Gly  
1 5 10 15  
Leu

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**